ISBN 978-82-326-3292-3 (printed ver.)
ISBN 978-82-326-3293-0 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

Luca Abele Piciaccia

Doctoral theses at NTNU, 2018:251

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Engineering
Department of Mechanical and Industrial
Engineering

NTNU

Doctoral thesis

Doctoral theses at NTNU, 2018:251

Luca Abele Piciaccia

Systems engineering for the subsea oil & gas industry – requirements elicitation through semantically aware technologies – a quantitative assessment

NTNU
Norwegian University of
Science and Technology

Luca Abele Piciaccia

Systems engineering for the subsea oil & gas industry – requirements elicitation through semantically aware technologies – a quantitative assessment

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2018

Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Engineering
Department of Mechanical and Industrial Engineering

ISBN 978-82-326-3292-3 (printed ver.)
ISBN 978-82-326-3293-0 (electronic ver.)
ISSN 1503-8181

IMT-report 2018:251

Doctoral theses at NTNU, 2018:251

Printed by NTNU Grafisk senter

Preface

This thesis is the result of a PhD project conducted at the Faculty of Engineering, Department of Mechanical and Industrial Engineering, at the Norwegian University of Science and Technology (NTNU). The work took place from 2012 to 2018.

The PhD project has been carried out in close collaboration with my main supervisor, Associate Professor Cecilia Haskins at the Department of Mechanical and Industrial Engineering, with co-supervisor Professor Stein Ove Erikstad at the Department of Marine Technology (NTNU). Essential to the work has been the cooperation with the Universita' di Roma Tor Vergata, Professor Roberto Basili and Assistant Professor Danilo Croce at the Department of Computer Science.

Industrial aspects and inter-university cooperation

A major Norwegian Subsea Contractor provides the dataset and partial funding.

University in Rome Tor Vergata - Department of Enterprise Engineering has been the university partner for this research.

The semantic technologies used in this research are made available by University of Roma, Tor Vergata in Italy through Reveal srl.

Reveal srl is a spin-off of the University, which in the last 20 years has developed, designed and widely applied human language technologies, semantic search platforms and Web mining tools for advanced applications, in close cooperation with the Department of Enterprise Engineering of the same University.

Trondheim,

15 June 2018 *Luca Abele Piciaccia*

Acknowledgements

I am indebted to several outstanding professionals and scholars who lent their support during the long months it took to complete this research journey.

First and foremost, my supervisor at NTNU, Professor Cecilia Haskins, without her guidance and assistance in ferrying the research from the industrial field to a proper academic domain this work would not have seen its conclusion.

Also at NTNU, Professor Hans Jørgen Dahl and Professor Stein Ove Erikstad provided inspiration and the proper foundation for some of the publications that became part of the work.

I am grateful to Professor Roberto Basili and Assistant Professor Danilo Croce from the University in Rome Tor Vergata, who were visionary enough to start with a pro-bono effort that paved the way for a long industrial cooperation in the then untapped subsea oil industry domain. Their world-class expertise in semantic technologies and commitment to this research constituted the bedrock for the quantitative results.

Ernst Petter Axselsen, PhD, as managing director of VEAS was so forward looking to embrace the artificial intelligence techniques applied in this research and set the conditions for their continued use across new domains.

Summary

From the year 2000 through to 2016, no floating production storage and offloading (FPSO) projects were delivered on time or on budget. This straightforward but astonishing reality highlights the challenges of requirements elicitation, management and eventual contract fulfilment within the oil and gas industry domain. This study is a step in the research for an improved method to elicit requirements in the subsea oil industry and apply it to project execution, with the aim to improve execution costs through significant cost-of-quality reduction.

Requirement elicitation is a challenge in every industry, but particularly acute for the subsea oil and gas extraction industry, this happens, is brought about by the short project award and execution times do not allow a thorough compilation of a requirement database from the multitude of contract documents and appendixes. Mainstream requirements management tools used in systems engineering in other domains generally cannot cope with the diversity and complexity of the subsea oil and gas tenders, the set of circumstances results in an ad hoc approach in an effort to counter the adverse effect of the incomplete information with the consequence of the inevitable requirements late discovery with bespoke and costly mitigations.

Contributions

The research has provided and quantitatively evaluated the following:

- A platform based on natural language processing in the subsea oil domain and its application through semantically aware technologies for the automatic establishment of a domain ontology.
- A semantically aware retrieval and browsing (SARB) system for the analyst, based on a domain specific AI generated ontology, validated by human experts.
- A system supporting the analysis of the technical project documentation through assisted review and decomposition of the available information also when scattered throughout the project documents.
- The quantitative confirmation of AI system capability of performing correct Inferences on document content in the subsea domain.
- The effectiveness evaluation of the inferences usefulness in supporting experts in the retrieval of relevant information for project documentation timely and exhaustive review.

Abbreviations

ASTM	American Section of the International Association for Testing Materials
BOK	Body of Knowledge
BoW	Bag of Words
ConOps	Concept of Operation
DNV	Det Norske Veritas
ERP	Enterprise Resource Planning
FPSO	floating production storage and offloading
INCOSE	International Council on Systems Engineering
ISO	International Organization for Standardization
LL	Lessons Learned
LSA	Latent Semantic Analysis
MACS	Material Classes
MLK	Machine Learning
NACE	National Association of Corrosion Engineers
NLP	Natural Language Processing
NSC	Norwegian Subsea Contractor
O&G	Oil and Gas
OM	Ontological Model
RA	Requirement a Analysis
SARB	Semantically Aware Retrieval and Browsing
SKOS	Simple Knowledge Organization Schemes
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TQP	Technical Qualification Program
VSM	Vector Space Model
XMT	Xmas tree

Glossary

This brief glossary is provided for disambiguating terms that can be found both inside and outside the domain covered by this text.

Artificial Intelligence

The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. (Oxford Dictionaries. Oxford University Press.2016)

Automata (sing. Automaton)

A machine or control mechanism designed to follow automatically a predetermined sequence of operations or respond to encoded instructions

Balanced scorecard

A composite scorecard involving a number of different measures or performance indicators.

Basis of Design

The document that describes the needs to be fulfilled by the EPC (engineering procurement and construction) contractor as understood by the client (operator of the oil field).

BoK

A set of knowledge within a profession or subject area which is generally agreed as both essential and generally known (Oliver 2012)

ConOps

A Concept of Operations (CONOPS) is a user-oriented document that "describes systems characteristics for a proposed system from a user's perspective. A CONOPS also describes the user organization, mission, and objectives from an integrated systems point of view and is used to communicate overall quantitative and qualitative system characteristics to stakeholders. (IEEE 1998)

Deep learning

Deep learning is a class of machine learning algorithms that:

- use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Successive layers use the output from the previous one as input.
- learn in d (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners.
- learn multiple levels of representations that correspond to different levels of abstraction; the levels form a hierarchy of concepts.

Design basis

The document that describes the solution that the EPC (engineering procurement and construction) contractor will realize to satisfy the company's (oil company operator of the oil field) needs as expressed by the company's "basis of design."

Enterprise resource planning (ERP)

The integrated management of core business processes, often in real-time and mediated by software and technology. ERP is usually referred to as a category of business-management software that an organization can use to collect, store, manage and interpret data from these many business activities.

Functional requirements

States what the system must do and trace to the functions that will accomplish them. They do not prescribe how the system will be built, only on what it shall perform and the relevant acceptance criteria.

IDEF0

The IDEF0 Functional Modelling method is designed to model the decisions, actions, and activities of an organization or system. It was derived from the established graphic modelling language Structured Analysis and Design Technique (SADT) developed by Douglas T. Ross and SofTech, Inc. In its original form, IDEF0 includes both a definition of a graphical modelling language (syntax and semantics) and a description of a comprehensive methodology for developing models. (www.idef.com)

Interface requirements

Specify input/output types, limits of flow, and timing at the interfaces between components. These requirements are increasingly important due to the supply chain globalization. Industry standards in interfaces are critical. Interfaces must be consistent for the system to work.

Machine learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. (Mitchell,1997)

Need

The original reason for introducing a requirement

Non-temporal performance requirements

Quantify expectations for properties of the system like cost, weight, size, power consumption, availability, security, etc.

Ontology

A set of concepts and categories in a subject area or domain that shows their properties and the relations between them.

Prescriptive Requirements

Predetermine a design choice. It is essential to ascertain with the applicable stakeholder if this requirement is meant to apply to the letter or whether it is a misstatement of a requirement in the form of design. Either an imposed constraint that will be followed or one of the other kinds of requirements

Requirement

A capability or attribute that must verifiably be met or possessed by a system or system component to satisfy a: contract, specification, standard, or other formally or legally applied governing document

Semantic Technologies

Semantic Technology defines and links data on a repository by developing languages to express rich, self-describing interrelations of data in a form that machines can process. Thus, machines also able to store, manage and retrieve information based on meaning and logical relationships. The core difference between Semantic Technology and other data technologies, the relational database for instance, is that it deals with the meaning rather than the structure of the data (https://ontotext.com/knowledgehub/fundamentals/semantic-web-technology/ 2015)

Semantics

The branch of linguistics and logic concerned with meaning. The two main areas are logical semantics, concerned with matters such as sense and reference and presupposition and implication, and lexical semantics, concerned with the analysis of word meanings and relations between them. (Oxford Dictionaries. Oxford University Press.2016)

SVM

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

Syntax

The arrangement of words and phrases to create well-formed sentences in a language. (Oxford Dictionaries. Oxford University Press.2016)_A set of rules for or an analysis of the syntax of a language ‘generative syntax’ also the branch of linguistics that deals with syntax.

Taxonomy

The practice and science of classification of things or concepts, including the principles that underlie such classification

Temporal performance requirements

Quantify the amount of time system has to respond to the stimulus (Response Time).

Xmas tree

used on offshore oil and gas fields, a subsea tree monitors and controls the production of a subsea well. A design taken from their above-ground cousins, subsea trees are sometimes called xmas trees because the devices can resemble a tree with decorations.

Contents

1 Introduction

Since the year 2000, the subsea oil industry struggles to keep up with the normative and technological fast-paced evolution. Products with a development life cycle measured in decades, as referred among many by Offshore Magazine already in 2006 have to comply with normative references being substantially reviewed and updated on average every second year. The normative standards influence heavily the product and its obsolescence and yet the industry does not seem able to efficiently implement a requirement management system with the productivity exhibited by the best in class provider of complex artefacts such as the space industry as pointed out by DNV-GL in 2014

This research addresses the causes and investigates the application of artificial intelligence (AI) to propose a viable solution able to expedite the review process, enabling the use of well-established requirement management systems in a timely fashion.

1.1 Background

From the year 2000 through 2016 no FPSO (floating production storage and offloading) projects were delivered on time or on budget, as found in the writings of Ree (2012). This straightforward but astonishing reality highlights the challenges of requirements elicitation, management and eventual contract fulfilment within the oil and gas industry domain. The industry is in dire need of research results for an improved method to elicit requirements in the subsea oil industry and apply it to project execution, with the aim to improve execution costs through significant cost-of-quality reduction (Erikstad 2014).

Requirement elicitation is a challenge in every industry, and the subsea oil and gas extraction industry is acutely affected by a shortcoming of effective solutions. This is the result of the short project award and execution times that do not allow a thorough compilation of a requirement database from the multitude of contract documents and appendixes, not fulfilling at least one of the tenets of systems engineering as laid out by the International Council on Systems Engineering (INCOSE). Mainstream requirements management tools used by systems engineering in other domains could generally cope with the diversity and complexity of the subsea oil and gas tenders, their use is hampered by the lack of a proper and exhaustive requirement collection and elicitation. This results in an ad hoc approach in an effort to counter the adverse effect of the incomplete information with the consequence of the inevitable requirements late discovery with bespoke and costly mitigations. In particular, sustainability needs and requirements, in both the domains of industrial lifecycle and industrial ecology appear not to receive the deserved attention.

Every requirement analysis stage relies on strong assumptions about the nature of the underlying information: terminology and normative statements in requirement documents (e.g. design specifications, standards) as well as components, devices and their parts involved in a system. However, these assumptions are fully implicit in the documents. They are for example labels, acronyms or names for the involved components or titles (i.e. column names) for tables or field names for records.

In other domains and scenarios, the study of handling large volumes of information has been already investigated and partially addressed. Recent years have been described as the era of big data as McKinsey reports (2017). For example, there are about 1 trillion web pages; one hour of video is uploaded to YouTube every second, amounting to 10 years of content every day, Walmart handles more than 10,000,000 transactions per hour and has databases containing more than 2.5 petabytes ($2.5x10^{15}$) of information, according to Alexa Internet (2017). The genomes of 1000s of people, each of which has a length of $3.8x10^{9}$ base pairs, have been sequenced by various laboratories, as reported by C.G van EL and the European Society for Human Genetics (2013)

This large volume of data requires automated methods of data analysis, which is what semantic technologies provide. In particular, machine learning, as this branch of artificial intelligence is called, provides a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future results, or to support decision-making process under uncertainty. For this reason, semantic technologies play a key role in many areas of science, finance and industry.

The industry domain experiences a similar degree of complexity even at the project inception phase. Large industrial plants have to conform to a vast array of regulations even before the specific process is implemented, as an example a look at the EPA effluent guidelines (2017) will enable the reader to assess the complex regulatory regime Compliance to the article of law has to be assured and documented, often over a collation of documents spanning decades. Supply chain and logistic requirements often shape the internal routines of a company and the implementation of ERP systems generally forces workflows to comply with a pre-defined pattern established a-priori from outside of the project/company.

Requirement management is one cornerstone of systems engineering, and when applied to large-scale organizations or systems of systems, it is easy to recognize the challenges as similar to those tackled by semantic technologies. Therefore, it is appropriate to consider the application of machine learning techniques to the system life cycle analysis as approached by system engineering, in order to automatize some of the different phases of the system development lifecycle.

1.2 The Research questions

The research has been guided by the following research questions:

RQ1:

Is it possible to build an effective and practically useful ontology or at least a taxonomy for an industry domain through semantic-assisted processes?

Does a bottom-up (semantically assisted) approach deliver better results than the top-down (axiomatic) method?

RQ2:

Can the ontology be applied to yield a near complete requirement register?

Is it possible to apply such ontological knowledge to extract the requirements embedded in natural language texts?

RQ3:

Can research demonstrate the value of this method for O&G

What is the optimal required level of semantic richness (ontology vs taxonomy) for a given development strategy (top-down or bottom-up)?

Is it possible to define and apply metrics that quantify the improvement of the overall project performance when requirements are extracted in this manner?

1.3 Contributions

The work to address these questions followed a stepwise progression resulting in four publications as follows.

1. Requirements Elicitation through Semantically Aware Techniques for the Subsea Oil Industry Systems Engineering INCOSE International Symposium 2014 Las Vegas May 2, 2014 - Wiley Online Library
2. Ontology-driven Semantic Search for Requirement Engineering INCOSE European Symposium 2014, Cape Town October 20, 2014 Wiley Online Library
3. Subsea oil production: The design basis INSIGHT JUNE 2O17 VOLUME 20 / ISSUE 2 - Wiley Online Library
4. The case for requirement management: How industrial design requirements are specified and executed in the Oil Subsea Production Systems (SPS) vs. Shipbuilding industries Society of Petroleum Engineers SPE-188583-MS 2017

The research questions and the resulting papers and contribution to knowledge are given in Table 1.

Table 1 The research questions, related papers, and contributions

RQ	Related Papers	Contribution
RQ1	1, 2	In these papers, a Semantically Oriented approach aiming to automatize relevant inferences involved in the Requirement Analysis (RA) stages is proposed. Natural language processing and machine learning techniques can be used to automatize semantic annotation of requirement documents by locating sentences expressing requirements and assigning them specific (ontological) types.
RQ2	1,2	This method is able to convey the semantic associations of the various terms returned and augments the awareness of the operator to classes of phenomena, since clustering of results indicates a cluster of results tightly connected to each other. The complete metadata set for each result is available the system returns the sentence that triggered the result and returns all available metadata for collection in export baskets enabling register compilation.
RQ3	3,4	"Subsea Oil Production, The Design Basis" by Piciaccia establishes through information models (model-based systems engineering) how the subsea production system design basis should be optimally developed. The MBSE integrates into this work by the theory of transactional processing, which is implicit in the universal thought process when one drafts behavioural models but rarely receives reference.
RQ1	2,3,4	The ontology has been built and validated, with a surprisingly good adherence of the SARB/SKOS results with domain experts' opinion. The use of the retrieval interface has been demonstrated and the multiple domain experts who reviewed the results do agree on the potential the technology has for augmenting the analysis capability of the engineering team.
RQ3	1,2	Baselines performance measures (with lowest baselines achieved with respect to deeper levels of the ontological hierarchy, i.e. at a finer grain) are significantly improved in all our tests. The relatively short time required to develop a semantic annotation tool for documents entering in the requirements analysis chain is a clear sanction of the potentials of these method in realistic operational RA settings.

1.4 Scope and Limitations

The research provided a demonstrator for the methodology and technologies where none existed in the subsea industry domain. Once the requirements were extracted, they were validated and found of sufficient quality to deserve further assessment on the most effective use of such application knowledge in the subsea oil industry. As part of the research plan, it is not envisaged to implement the findings in a future project until the measure of effectiveness indicated in the section 2 are satisfied.

1.5 Ethical issues

No ethical issues have been identified for the scope of the research. The research domain can be described as the application of Artificial Intelligence to the industrial Oil production industry. The work is analytic in nature and does not entail experiments involving animals, personnel or general public participation.

Since the inception of this research, a societal debate is ongoing on the more general ethical implication of AI in the workplace, seen by some as a threat able to replace paid workers with machines. This research and its foreseeable application is intrinsically dedicated to opening new analysis possibilities, not automatizing the existing ones, therefore, also these emergent ethical concerns do not apply to this specific work.

1.6 Structure of the thesis

The thesis has two main parts; Part I: main report and Part II: articles.

Part 1

- Chapter 1 provides the premise and background to this project.
- Chapter 2 provides a first introduction to the research design and the semantic technologies involved to support the requirement analysis in the subsea domain.
- Chapter 3 frames the research in the domain within a systems engineering context.
- Chapter 4 reviews the application of machine learning techniques as relevant to the requirement analysis phase and presents the ontological model defined in order to represent the information targeted in the context of this project.
- Chapter 5 examines and evaluates the application of the chosen Machine Learning techniques and developed prototype to real scenarios;
- Finally, chapter 6 discusses the outcome of this work as well as the possible directions.

Part 2

- Published articles
- Appendix 1 : Physical Quantities
- Appendix 2 : The MAC classes

2 Research Design

This selection describes the case study research, the rationale for the approach and an evaluation of the method applicability following the argument laid out by Flyvbjerg in 2006. The requirement extraction from natural language texts is approached through case studies, where the well-known support vector machine (SVM) technique is combined with natural language processing (NLP) to construct and make available a semantically aware retrieval and browsing (SARB) to support the analyst in identifying instances semantically connected to the query in the application domain. The technique is then applied to specific documents whose instances supply cases for the falsification or validation of the stated hypothesis.

2.1 Use of case study methods

A classical misunderstanding about case study research concerns generalization or the presumption implying that a valid generalization is impossible or at least so deeply flawed to render it invalid in the scientific domain. The old approach to case studies would dismiss its use for generalization purposes, here follows an example of such criticism:

> Case Study. *The detailed examination of a single example of a class of phenomena, a case study cannot provide reliable information about the broader class, but it may be useful in the preliminary stages of an investigation since it provides hypotheses, which may be systematically tested with a larger number of cases (Abercrombie et al. 1984: 34).*

The underlying assumption of this criticism is that without a clear and pre-constructed rules framework to be tested through extensive statistical methods applied to a large sample population no generalization is possible. To this viewpoint Flyvberg counters

> *It is not that rule-based knowledge should be discounted: it is important in every area and especially to novices" [case studies are] "important for the development of a nuanced view of reality, including the view that human behavior cannot be meaningfully understood as simply the rule governed acts found at the lowest levels of the learning process, and in much theory (Flyvberg 2006: 6).*

This is of particular significance to the research since NLP is analysing language and language is inextricably linked to human behaviour. By providing analysis and context to semantic phenomena, we are in fact enabling the deeper understanding of the domain, its constituents and the relationships among them as practiced by humans through their behaviour expressed by words.

Eckstein, contravening the conventional wisdom in this area, goes so far as to argue that case studies are better for testing hypotheses than for producing them. Case studies, Eckstein asserts, are "… most valuable at that stage of theory building where least value is generally attached to them: the stage at which candidate theories are tested." (Eckstein, 1975: 80).

The research will substantiate its claim to legitimacy through expert validation of the results and provide sufficient corroboration of its general value through quantitative measures and the underlying linear algebra theory, which provides a sound and rigorous foundation for the whole study.

For the research an accurate case selection was done according to the principles suggested by by Flyvberg (2006):

Information-oriented selection To maximize the utility of information from small samples and single cases. Cases are selected on the basis of expectations about their information content.

Critical cases: To achieve information which permits logical deductions of the type, 'if this is (not) valid for this case, then it applies to all (no) cases.'

Paradigmatic cases: To develop a metaphor or establish a school for the domain which the case concerns.

On the matter of subjective bias in case studies, this research departs from the old school approach of a top-down ontology. The combination of SVM and NLP provides a true bottom-up ontology extracted from the BOK that characterizes the domain. This extraction is free from the bias intrinsic in a top-down approach where, by necessity, the human experts trying to compile the ontology will unwittingly insert their bias and omit the knowledge they do not possess. NLP and SVM are also able to expand the ontology and the semantic connection in a dynamic way through the addition of new documents and deletion of obsolete ones to and from the BOK as knowledge evolves.

> *It must again be emphasized that despite the difficulty or undesirability in summarizing case studies, the case study method in general can certainly contribute to the cumulative development of knowledge. (Flyvberg, 2006: 25)*

2.2 Case study overview

For the remainder of this research we consider as a case the application of semantic technologies to a specific problem that can be placed in a broader context for later generalization and where the initial conditions (inputs) can be defined adequately. As an example, we use the automatic supervised bottom-up extraction of the domain ontology.

A major Norwegian subsea contractor (NSC) provided examples of archived large contracts, lessons learned records and a wide range of technical specification, and granted permission to use this material to integrate public domain information collected in digital form from several internet specialized sites. This wide array of digital text is processed to extract the ontology based on the real language usage in the domain of interest. The result is then validated through "oracles" i.e. domain experts with long experience. This "case" demonstrates the effective capability of the technology and can be scaled throughout the domain or even across domains.

2.3 Progress plan

The path followed by this research is illustrated in figure 2.1 and can be summarized as follows: The first stage is the construction and validation of a subsea domain ontology, in itself a research activity. Additional introductory material on the creation of an ontology can be found in the works of Noy and McGuinness (2001), Manning, Raghay and Schuetze (2008), Velardi, Missikoff and Basili (2001). The ontology is then applied via SARB techniques to the body of project documentation for a selected project. The aim is to elicit all the requirements present in the project documentation and compile the relevant requirement register. Requirements are allocated to entities (artefacts) as needed in a 1-to-1 or 1-to-many relationship. At this point the products and system documentation from the supplier undergoes the same process, to elicit the properties of each piece of equipment associated with the project scope of work.

Then, the manually prepared, real-life project documentation status is compared with the register to identify the requirements that are detected by the semantic system and not detected by the standard business process and those that the semantic system did not detect (if any). The system is able to load the textual sources as they are currently stored and managed within the information systems without pre-processing, extract and index documents according to a number of semantic phenomena, use these indexes to support complex forms of question answering and finally support concept-driven browsing as an exploratory practice. In its first deployment, a community of expert/specialists, in a distributed intranet based environment, will use the system. The system will reuse a number of open source technologies whenever applicable.

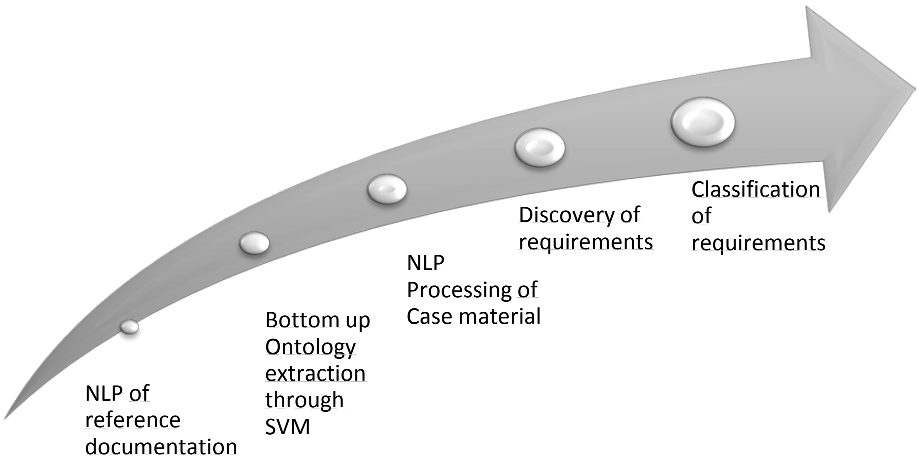


Figure 2.1 Towards requirements extraction through semantic technologies as applied in this research.

The work schedule has been organized around the following stages, each achieving a verifiable result and building on the previous phase in a sequential activity schedule. The stages are illustrated in figure 2.2.

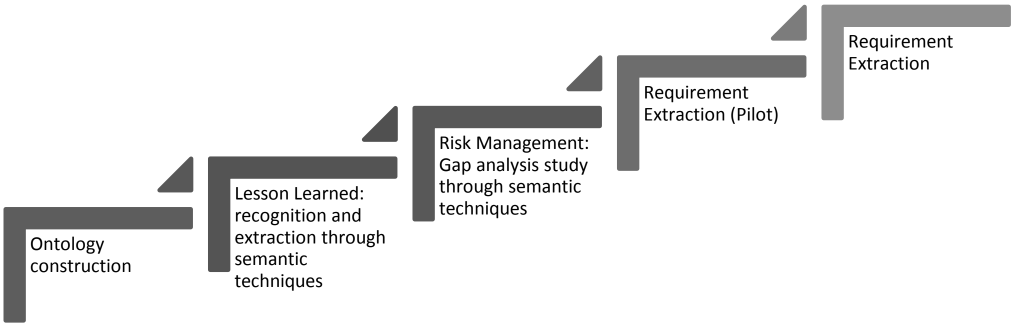


Figure 2.2 The intended progression from the basis of NLP to requirements extraction

The methods, applied in other domains to automatic supervised ontology extraction, have exhibited a yield of over 97% correctly identified terms after review by domain experts. This benchmark is used to assess the suitability of the approach to the domain under research.

2.3.1 Preliminary Stage: Literature review, state-of-research

The initial research activity has been spurred by the acute need to find a solution to the overwhelming number of requirements that the researcher had to manage as engineering manager of major subsea projects. After years of successfully applying systems engineering techniques, the rapidly expanding number of disparate regulations caused the engineering management function to become a bottleneck in project execution. After a long search for a suitable support technique, a presentation by prof. Basili at the 2012 INCOSE symposium in Rome offered the hint of a possible development through semantic technologies applied to industrial endeavours.

After an extensive review of the theory of design and system engineering to properly frame the research starting point, an intensive period of exploration into the ontologies and description logic combined with machine learning in semantic search followed. Since the research was oriented in the application of new and complex technologies to a necessarily limited number of cases, the most promising research approach that displayed the necessary theoretical foundations whilst being applicable to a limited number of direct observations was considered for the case study research.

2.3.2 Stage 1: Ontology construction

As foundation for all subsequent work, it has been necessary to establish an ontology of the domain where the research would be performed. For the scope of this research, ontology is defined as "a set of concepts and categories in a subject area or domain that shows their properties and the relations between them" (Oxford dictionaries 2016).

In this stage, a specific domain ontology is extracted through automatic supervised natural language processing NLP algorithms and structured through support vector machine SVM processing. The novelty of this research ontology consists in its automatic creation from large datasets (bottom-up) as opposed to the application of a pre-constructed taxonomy to the domain. In the years 2010-16 attempts at creating top-down ontologies were performed by Norwegian oil companies and consulting bodies such as DNV GL (e.g., DNVGL-RP-O101). A wide application of such ontologies has not been successful. One of the contributing factors of this limited success is the extreme difficulty of transposing a pre-defined ontology to new industrial realities where conventions on concepts and their relations already exist but are not formally explicit since it is a daunting task to identify them without machine learning support.

The subsea domain draws upon a body of knowledge (BoK) constituted of the main internationally recognized standards for the domain i.e. ISO 13628 vol. 1 to 17, API 6A to 6D (as included in ISO 10423:2009). In order to represent the concepts usage in practical applications, the BoK was integrated with the text of several contracts provided by the major Norwegian subsea contractor (NSC) complete with their extensive technical sections. In order to further reduce bias and expand the BoK, several public domain relevant published articles were included. Figure 2.3 illustrates the hierarchical prioritization of these documents.

NLP was applied to the BoK. After the addition of each document, the ranking of the terms was automatically updated to assess the contribution of each document to the overall ontology. Semantic distances were computed among all terms through the SVM algorithm. The achievement of a self-supporting critical mass of lexical instances, capable of integrating further additions in a stable ontology was considered confirmed when the addition of new documents did not significantly alter the SVM provided classification and semantic distances. It is to be noted that the machine learning algorithms continuously update the classification at each data source addition or removal.

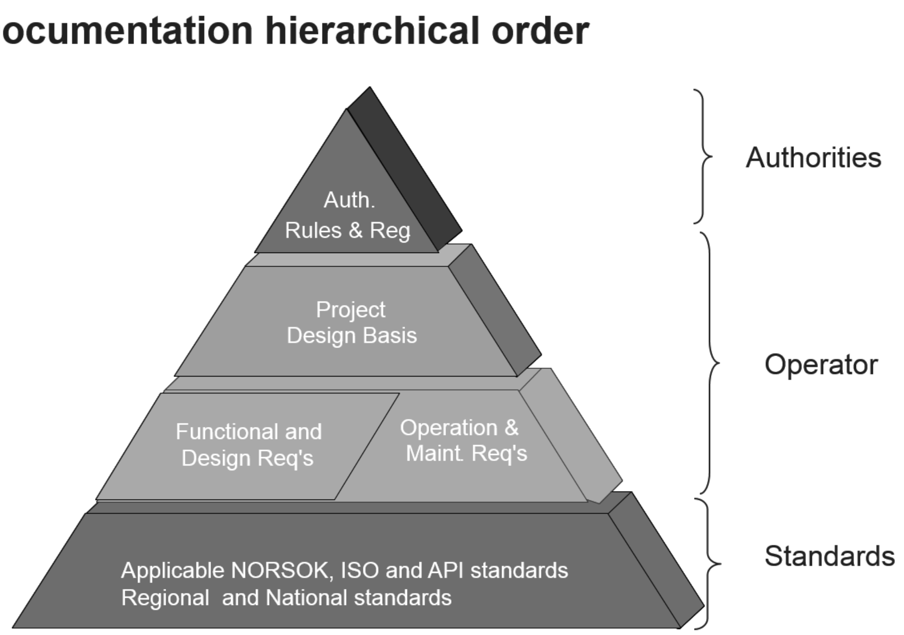


Figure 2.3 the order of prevalence among the different classes of documents in the subsea oil industry; the content of an upper class document overrides the lower class in case of interpretation conflicts

The result of this first stage was a list of around 2000 automatically identified relevant terms, which fulfilled a set of stringent rules for eligibility in the "most relevant" concept list. Competent professionals with extensive work experience in the domain were asked to verify the result. The applied methods are considered fit for purpose if over 97% of the terms have been correctly recognized as legitimate members of the ontology in the narrowly defined domain and as being in current use in the subsea oil community. This stage outcome satisfies the modelling requirements within kernel-based learning, as described by Chen, Lin and Shoelkopf (2003).

2.3.3 Stage 2: Lesson Learned: recognition and extraction through semantic techniques

Once a reliable and verified ontology was established, it was necessary to prove its applicability to the subsea oil domain. Through the SVM method, as already validated by the work of Manning, Raghavan and Schuetze (2008), a SARB approach to information retrieval was successfully attempted. The method was applied to the NSC subsea lessons learned database, to prove the feasibility of semantically assisted information retrieval in the subsea oil domain. The method exploited the semantic distance ranking to infer correlation among text sections residing in separate records of a large unstructured database to identify and extract the sections semantically closest to the text originating the query. The semantically assisted retrieval demonstrates significant improvement over even an advanced lexical search (i.e. Google style) as it was able to retrieve related concepts, e.g. a query on fasteners will retrieve screw, rivets, bolts, etc.

The objective is to implement a highly efficient, AI assisted, lessons learned (LL) retrieval tool arranged in a web-based service. The choice of a web-based solution provides the utmost scalability and flexibility with virtually unlimited access for all authorized users of the system. The main industrial advantage of this system will be the lowered threshold for lessons learned archival and retrieval since the SARB solution is format independent thereby affording the capability to exploit the entire historical LL library without any reformatting. Avoiding the reformatting allows the organization to dispense with a well-known source of error that has the added disadvantage of requiring resources without adding intrinsic value to the dataset or the project.

The approach is conceptually applicable to all natural language information repositories, of particular interest the collation of field reports, health and safety notifications, equipment recall notes, and maintenance records. In these domains, the assistance provided by SARB can be invaluable for a root cause analysis of phenomena with low frequency, possibly spread over several locations, whose reporting is performed across disjoined collection systems. The ability of SARB to quickly process raw data and suggest their relevance based only on content can be a game changer in these applications. Clusters of reports can be autonomously identified and brought to the attention of the specialist for further assessment.

2.3.4 Stage 3: Risk Management: Gap analysis study through semantic techniques

Technical documentation is in an ever-evolving state. Technical specifications change over time with considerable speed and in a typical project of 2 years duration one is almost assured of having to face at least one revision of the critical documentation. These techniques can be applied to the analysis of evolutionary versions of technical requirements documents, to further validate the semantically assisted classification and recognition of "hard" requirements, i.e. requirements directly linked to physical properties, such as pressure, temperature, and parameters expressed by quantitative measures. Of particular interest is the semantic analysis of the location in the document where specific topics are treated. Requirements insisting on an artefact are usually not concentrated in only one section of the document, but can be found almost at random locations throughout the text, and considerably complicate the analyst's work. Stage 3 results in an "anchored versioning" tool able to identify evolutions across different versions of the same document regardless of its length and, more importantly, identify disjoined areas of the document covering the same topic. This feature is most important when dealing with nested references within an overarching document. The research intention is to achieve a tool that, through the application of the SARB algorithm to both versions of the document and collating the results through a browser-based graphical use interface GUI, enables the analyst to simultaneously see both versions, be presented with the text found as correlated by the SARB analysis, and "jump" directly back and forth across the document to the relevant sections for a full comparison.

2.3.5 Stage 4: Requirement Extraction (Pilot)

The method is applied in the analysis of project technical requirements documents, to further expand the semantically assisted classification and recognition of requirements including the "soft" ones, i.e. those expressing properties not characterized directly by a quantitative measure (e.g. interchangeability, maintainability, and ease of use). Quantitative measures of accuracy and performance indicators are provided to quantitatively asses the performance of the system and the improvements the system will deliver.

2.3.6 Stage 5: Requirement Extraction

Once stage 4 delivers stable and predictable results, the method is applied to the real life project from the NSC for extracting the requirement register. Quantitative measures of accuracy and performance indicators are established beforehand to quantitatively assess the improvements the system will deliver. Tracking the analysis over time will provide a measure of effectiveness for the process. A balanced scorecard or similar quantitative approach is used to quantify the effective impact the activities left uncovered and estimating the effect of the delay in uncovering them in the overall project performance.

3 Systems engineering in the subsea domain

3.1 Systems engineering foundation of the research

In this work the definition of Systems Engineering as laid out by INCOSE has been adopted:

> *Systems engineering is an engineering discipline whose responsibility is creating and executing an interdisciplinary process to ensure that the customer and stakeholder's needs are satisfied in a high quality, trustworthy, cost efficient and schedule compliant manner throughout a system's entire life cycle. (INCOSE, 2017)*

The systems engineering process is usually comprised of the following seven tasks: State the problem, **i**nvestigate alternatives, model the system, integrate, launch the system, assess performance, and **r**e-evaluate. These functions can be summarized with the acronym SIMILAR. This Systems Engineering Process is shown in Figure 3.1. It is important to note that the Systems Engineering Process is not necessarily sequential. The functions are performed in a parallel and iterative manner." as formalized by Bahill and Gissing.

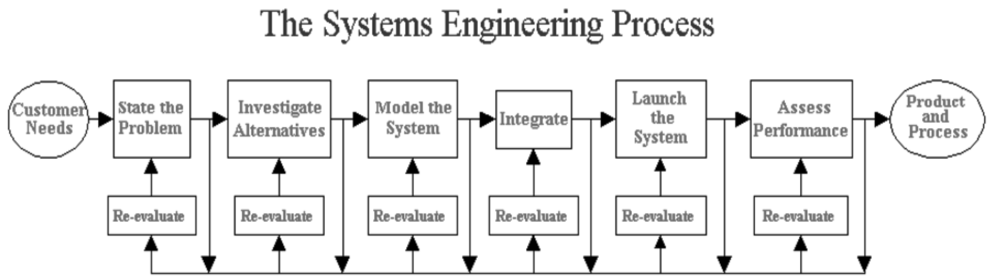


Figure 3.1. The Systems Engineering Process as formalized by Bahill and Gissing (1998).

The system engineering approach is embodied in the oil and gas industry by industrial processes that assign specific responsibility for the execution of each phase to resources in the company organization. An example of such implementation is given below in Figure 3.2 where an IDEF0 inspired representation of business process section is depicted (NSC proprietary information is removed).

The consensus among INCOSE fellows is that a system is a construct or collection of different elements that together produce results not obtainable by the elements alone. The needs to be fulfilled are the reason for crafting a system. The performance measure for this dimension has to be objective and quantitative. Therefore, a high-level need such as "the system shall operate control the flow of oil from the subsea well to the platform" has to be expressed through statements whose compliance can be verified, i.e. requirements containing more details. In SE, the term validation is often summarized as "build the right system", i.e. a system that fulfils all the needs, and the verification as the activity that ensures compliance with all the requirements we used to express our needs. Requirements are actively elicited and collated in a requirement register.

Requirements elicitation appears simple, but it is the most challenging activity in every project, and because it lays down the foundation for all the subsequent work, it has the widest ranging consequences. Whilst needs can be broadly expressed at a high level, requirements demand a specificity which often is not to be found in the early statement of the problem. Often the project objectives are spread across several documents and in multiple places within each document with very few quantitative measures. Different authors express their idea of the final result in a variety of ways and may on one hand fail to be explicit about what they consider trivial and taken for granted and on the other hand rely on blanket coverage of the requirements by referring to standards. This latter approach is the norm in the O&G industry and carries with it the unfortunate proliferation of implicit and conflicting requirements that are invariably discovered as such during the advanced stages of the project, where they have the potential to induce late changes and the highest negative impact on schedule and cost.

3.2 Requirements in the subsea domain

The current state of the art in the subsea domain uses lengthy documents compiled in natural language and purporting numerous references to additional texts and appendixes to describe the scope of work for the contract being awarded. This results in an intricate description where lack of clarity is rife and the very nature of natural language inhibits an efficient individuation of the requirements, requiring specific techniques to unravel the intrinsic complexity as discussed by Giannone, et al. (2011). "Late requirements" i.e. requirements that surface only during a further detailed review of the specifications, are a major cost-of-quality driver for the subsea industry, significantly deteriorating operative margins.

In the subsea domain, now a mature industrial field, oil extraction is driven towards deeper and more challenging production areas, up to and including the artic. Technology development has continuously to keep pace with this evolution. A significant challenge is posed by the qualification of a new technology as illustrated in DNV-GL (2011) where the correct and exhaustive description of both the operating environment and concept of operation (ConOps) as exemplified by INCOSE (2015) and ISO/IEC 15288 (2008, 2015) is the pillar for the subsequent activities but has to be collated from several disparate sources through documents written by engineers with different backgrounds. Depending on the development model, requirements capture may be performed nominally once near the beginning of the development cycle, or as for agile methods, be a continuous activity, see Blanchard (1991). The reason for eliciting requirements is the same, understand the needs of the stakeholders well enough to support the architecture design process and deliver a "fit for purpose" system (INCOSE 2015). Yasseri (2014) offers a detailed analysis of systems engineering applied to subsea developments, which is applicable, with the necessary adjustments, to the vast majority of offshore oil fields. Subsea projects need to fulfil concurrent delivery schedules, integrating the work of large numbers of technical personnel of different background and nationality, allocating requirements across several interfacing system elements, acquired through long supply chains spanning literally across continents. Systems engineering (SE) is proving itself as a main enabler of subsea projects and the relevant value chain.

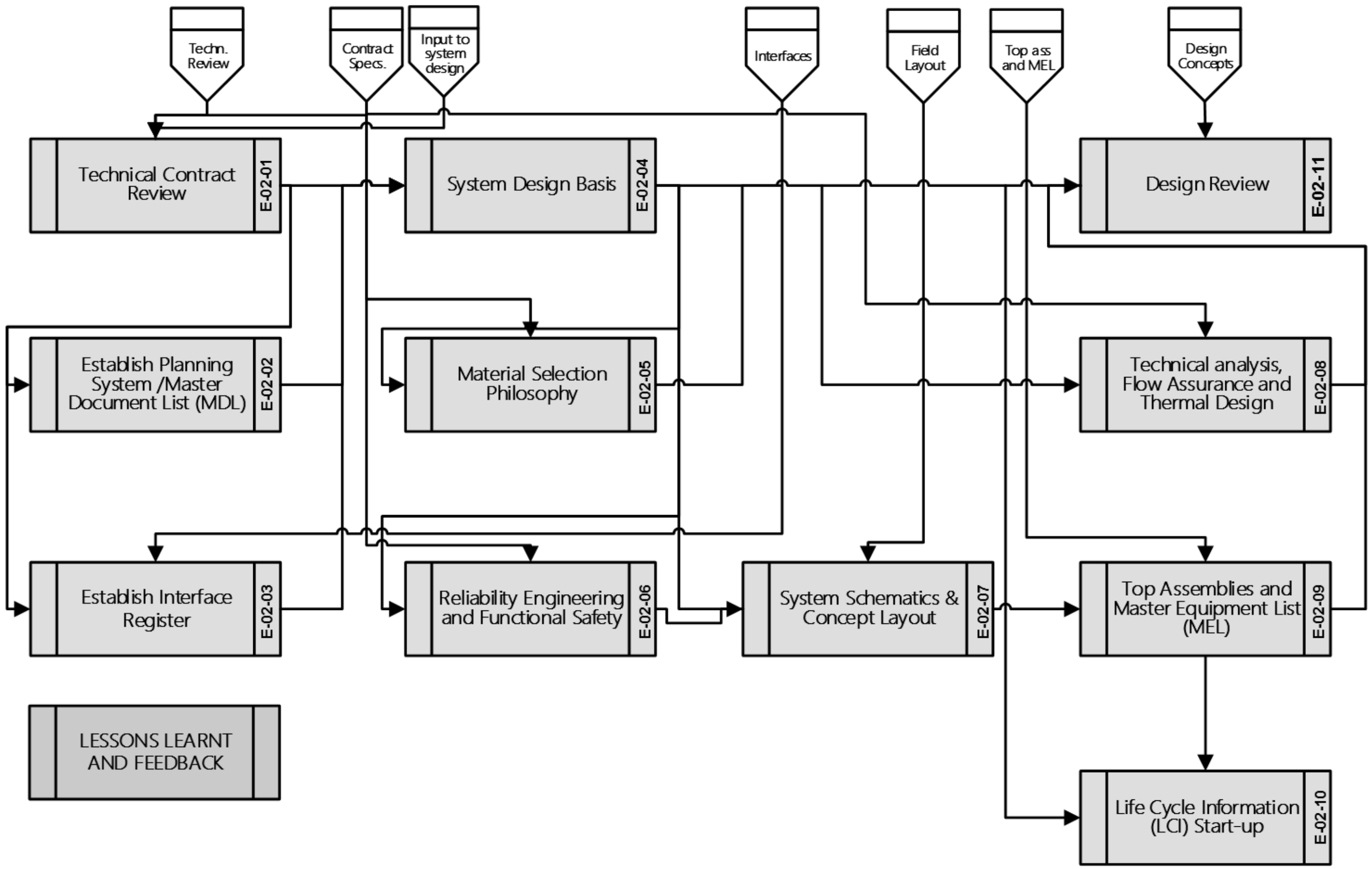


Figure 3.2: The implementation of selected systems engineering functions in a major subsea oil industry supplier organization

This research originated from the observed gap between the best practices implemented in the most advanced industries, such as automotive, and the oil industry. Following the approach suggested by the INCOSE Systems Engineering Handbook (INCOSE, 2015), it is possible to identify eight critical areas to be addressed for the successful completion of a project, as shown in figure 3.3.

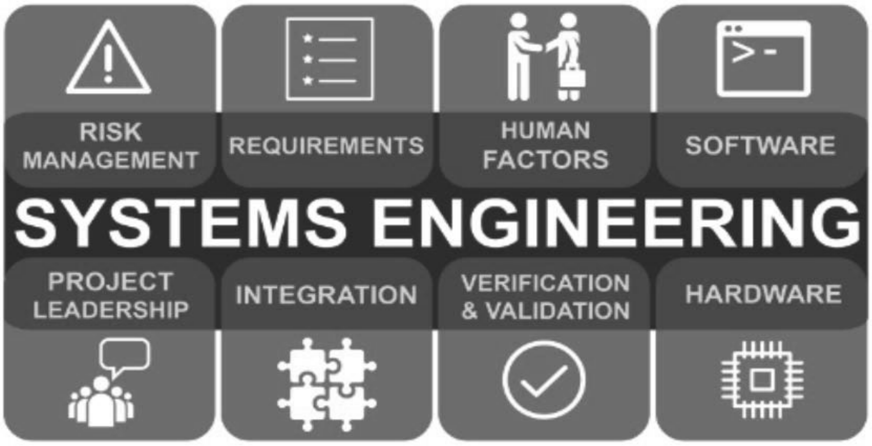


Figure 3.3: systems engineering aspects according to INCOSE (https://www.incose.org/systems-engineering)

The contribution intended for systems engineering is centred on **requirements** discovery, since the semantically assisted requirements identification has the capability to significantly enhance both the quality and completeness of the requirements collation and analysis. Automatic supervised requirement discovery has the potential of cross-fertilization among technical disciplines, since it proposes relevant finds even from specific domains not part of the supervising team experience, exposing the supervisors to a wider results area and the possibility of discovering associations not immediately evident. Beyond this direct impact, the enabling of efficient requirements management positively affects **risk management**, since the adverse event implicit in the definition of risk is very often the consequence of failing to satisfy a “forgotten” or “invisible” requirement. In addition, **verification and validation** are dependent on the explicit formulation of needs through requirements. Bringing to the surface associations among and across statements of needs whilst explicating the links between needs and requirements through semantic distance enables the discovery of “orphan” requirements, e.g. requirements carried over (probably from previous projects) which have no anchoring in any present needs, together with duplicates or unaddressed or under-addressed needs.. Better **project leadership** is enabled by commercially available requirements management systems. Requirement management systems cannot deal with natural language documents very well. Without an automatic supervised system requirements must be manually evaluated, an activity that ties up large amounts of expert resources and often cannot be concluded on time. A more efficient and integrated **hardware and software** solution, capable of interacting with **humans** in a seamless way, is possible when all the needs of the man-machine interface are explicated and implemented. The concept of operations can then properly consider the fallibility and uncertainty added when humans are in the loop, as well as making full use of the extraordinary adaptability that humans exhibit on a variety of problem solving tasks.

4 Machine learning for requirement analysis

This section provides an introduction to machine learning as a branch of AI to enable the design and implementation of semantic technologies for system engineering. In particular, it will focus on learning algorithms behind the selected semantic technologies used in the context of the current project. This section is written in close cooperation with Professor Roberto Basili and Assistant Professor Danilo Croce, University of Rome Tor Vergata , Faculty of Engineering, University of Roma, Tor Vergata.

4.1 Machine learning for semantic technologies

Since the 1990s, attention has been placed on empirical and statistical methods that were well founded on the information theory, introduced by Shannon (1948), and quite popular for natural language processing tasks, such as machine translation, before the 1960s. In the AI community, a significant shift was observed; from manually constructing grammars to partially or totally automating the acquisition of such information, through statistical learning methods trained on large annotated, or not annotated, natural language collections of linguistic information, i.e. document corpora (Cardie and Mooney, 1999)

These empirical approaches were based on the idea that people's language capability is not entirely dependent on cognitive abilities present in the human brain. An initial structure is assumed, which causes preference for certain ways of organizing and generalizing from sensory inputs over others, as no learning is possible from a completely blank slate, a *tabula rasa*. Nevertheless, empiricist approaches assume that the mind does not begin with detailed sets of principles and procedures specific to the various components of language and other cognitive domains as, for instance, theories of morphological or syntactic structures. It is rather assumed that a baby's brain begins with general operations for association, pattern recognition, and generalization, and that these can be applied to the rich sensory input available to the child to learn the detailed structure of natural language (Manning and Schütze, 1999)

An empiricist approach to *Natural Language Processing* enables the application of semantic technologies to language and it suggests that it is possible to learn the complicated and extensive structure of language by specifying an appropriate general language model, and then inducing the values of parameters by applying statistical, pattern recognition, and machine learning methods to a large amount of language in context. The popularity of such approaches grew after the initial success obtained for the statistical processing of low-level language problems, such as those related with speech recognition and syntactic tagging (Abney, 1996). Ever since the early 1990s, the choice of automatically processing such massive quantities of free text (commonly referred to data-intensive approach or corpus-based approach) has contributed to the development of a number of methods and techniques with an application to a great variety of natural language acquisition and understanding problems, including: automatic extraction of lexical knowledge, lexical and structural disambiguation (part of speech tagging, word sense disambiguation and prepositional phrase attachment disambiguation), grammatical inference and robust parsing, information extraction and retrieval, automatic summarization or machine translation.

4.1.1 Statistical learning methods

Statistical learning methods make the assumption that lexical or grammatical observations are useful hints for modelling different semantic inferences, in this context, linguistic observations provide features for a learning method that are generalized into predictive components in the final model, induced from the training examples. Mitchell (1997:3) provided a ground breaking definition of a learning program:

A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.

The Semantic Analytics Group, part of the Artificial Intelligence Group at the University of Roma Tor Vergata offers a clear definition of machine learning when dealing with language.

In Natural Language Processing (NLP) such formulation supports the following definitions for learning systems that can be applied to software engineering.

- **T** represents a linguistic task, usually an interpretation, such as in semantic annotation or document classification tasks. In this study, such semantic processing task will be formulated as a statistical classification problem: the target is to identify the sub-population to which new data belong, where the identity of the sub-population is unknown (the test data), on the basis of a training set of data containing observations whose sub-population is known (the training data). For example, in the document classification task, texts are mapped to a set of classes that characterize the document topics, e.g. a document refers to sport, economics or politics. The objective is the acquisition (from data) of a function $y = f(x)$ that is able to associate each text x to is corresponding class y.
- **P** represents the performance, thus measuring the quality of the resulting interpretation power. It depends on the task objectives and the learning system requirements. For example, if one is interested in the quality of a document classification system, the accuracy score can be employed as the percentage of correctly classified texts. However, if the learning algorithm improves the performance according to other aspects, e.g. the time needed for classification or the resource requirements of the produced learning system, other performance measures can be employed.
- **E** is represented by data as observations available about the target task. The idea is that a learning system exploits such information in order to acquire competences to resolve the target problem; the more information observed, the better is the performance P to solve the task T. In the classification task, experience is provided by the documents themselves that are examples of *x* providing different aspects of the target problem in terms of linguistic observations, such as lexical, grammatical or syntactic information.

4.1.2 Support Vector Machine

Different machine learning algorithms exist in order to exploit data evidence and acquire a model of the target task, as discussed in Bishop (2006). Leveraoing on ears of experience in the application of several ML technologies , the Support Vector Machine (SVM) learning algorithm, as discussed in (Vapnik, 1998) and (Basili & Moschitti, 2005), will be employed as it provides an effective learning paradigm to satisfy the objectives. SVMs can be thought of as methods for constructing classifiers with theoretical guarantees of good predictive performance in terms of the quality of classification on unseen data. The theoretical foundation of this method is given by statistical learning (Vapnik, 1998).

Formally, the goal of a statistical learning algorithm is to learn a mapping from inputs x to outputs y, where $y \in \{1, \dots, C\}$, with C being the number of classes. If $C = 2$, this is called **binary classification** (assuming $y \in \{0, 1\}$). If $C > 2$, this is called **multiclass classification**. One way to formalize the problem is as **function approximation**. Assume $y = f(\mathbf{x})$ for some unknown function f, and the goal of learning is to estimate the function f given a labeled training set, and then to make predictions using $\hat{y} = \widehat{f(\mathbf{x})}$, a function estimation. The main goal is to make predictions on novel inputs, meaning ones not seen before (this is called **generalization**), since predicting the response on the training set is trivial. This requires data to construct prediction rules, often a lot of it. Now suppose an available set of measurements (x_i, y_i) or (x_i, g_i), i = 1, …, N, known as the training data, with which to construct our prediction rule. This is applied to the creation of an ontology that models oil extraction within the subsea domain. In the modelling, each concept within this ontology represents a class. Sentences are linguistic objects modelled through artificial representation **x**. An example of binary classification is the recognition of whether a sentence does or does not represent a requirement. Examples of multiclass classification are the classification and assignment of sentences expressing requirements to the concept of the ontology.

4.2 SVMs for statistical machine learning

Support Vector Machines (Vapnik, 1998; Basili & Moschitti, 2005) represent a classifier belonging to the family of kernel methods (Shawe-Taylor and Cristianini, 2004), which are a large class of learning algorithms based on inner product vector spaces. SVM is one of the best known learning algorithms based on statistical learning theory, and has proved to outperform many other categorization algorithms (Wa'el Musa Hadi, 2007). In addition, SVM are able to handle large vector spaces with excellent accuracy (Xue Li, 2005). Classifying data is a common task in machine learning, and with respect to systems engineering, it enables classifying requirements according to a discrete score when requirements are sentences represented by points in a multi-dimensional space that can be assigned to a class in consistent with their type or capability. A clarifying example of the application of the concept can be found in the Article 4 "Ontology-driven Semantic Search for Requirement Engineering", where the "2-d plot of the semantic query "contraction"." illustrates the processing in a visual representation where "The distance among the text boxes in the graph is proportional with the semantic distance among the terms contained by each box".

In a supervised-learning perspective, a training set of data points is given, each of which belongs to one or more classes., The goal is to predict the class for a new data point from a disjoint testing set that can be assigned on the basis of prior learning from the given training set. SVM is the method employed in this project for the classification of requirements among given classes of ontological types or capabilities. First suggested by Vapnik in the 1960s, SVMs are closely related to neural networks and they perform classification by constructing an m-dimensional hyper-plane that optimally separates data into two classes (classical formulation) as well as into multiple classes. SVMs rely on kernel functions and, depending on these, they are an alternative training method for polynomial, gaussian functions and multi-layer perceptron classifiers (like sigmoid) in which the weights of the network are found by solving a quadratic programming (QP) problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem.

In a common binary classification task, SVMs look for the optimal separating hyper-plane between the two classes by maximizing the margin between the classes' closest points. Points lying on the boundaries are called *support vectors*, and the middle of the margin is the optimal separating hyper-plane. The classification task has been formulated as a quadratic optimization problem (Vapnik, 1998), which can be solved by known techniques, employing Lagrangians, and this is the problem solved by the SVMs. Further, it can easily be extended to *k*-class classification, for example by constructing k two-class classifiers (Rifkin & Klautau, 2004). Thus, considering the linear separable case, being:

- X The set of training data points, for which classes are known,
- $x \in X$, a data point,
- w a weight vector, normal to hyperplane

Then following function defines a separating hyperplane:

$$\sum_{i=1}^{m} w_i x_i + b = w^T x + b = 0$$

And, the decision to establish the class of a new point is given from the sign of the function. That is, the classifier, based on an inner product is:

$$f(x) = sign(w^T x + b) = \begin{cases} +1 \\ -1 \end{cases}$$

A linear classifier may not be the most suitable or effective approach for discriminating the classes. When a linear separator cannot be found, data points are projected into a higher-dimensional space where the data points become linearly separable. Such projection can be realized through the so called ***Kernel Tricks***, and the inner product in the classifier function can be computed in the new space without an explicit mapping, thereby avoiding time and memory consumption normally associated with the high space dimensionality of the new space. In the classification phase, the hyper-plane is not directly defined, as it is the linear combination of SVs, but the classification is still feasible in terms of the similarity (the dot-product) among the novel instances and the support vectors. The explicit representation of the novel feature space is never built and therefore it is called implicit feature space. The learning algorithm will select the most representative instances and characteristics in the implicit space, i.e. the space dimensions. Such methods provide effective statistical predictions defining meaningful similarity (i.e. kernels) functions among examples. It allows defining a learning paradigm whereas the algorithm can be directly applied over complex linguistic structure. These effective models can be acquired without focusing the attention over artificial representations and a novel and more effective representation can be defined in the implicit projection space.

The analyst has to take into consideration the higher degree of adaptability provided by Kernel based methods, which might induce convergence problems in the response to some particular instances of "noise" in the data. Kernel methods are beneficial because the combination of kernel functions can be integrated into the SVM, as these are still kernels, as discussed in (Shawe-Taylor and Cristianini, 2004). Consequently, the kernel combination

$$\alpha K_1 + \beta K_2 + \gamma K_3$$

linearly combines in a disjunctive form the contribution of kernel K_1, K_2 and K_3. Here, parameters α, β and γ weight the combination of the three kernels.

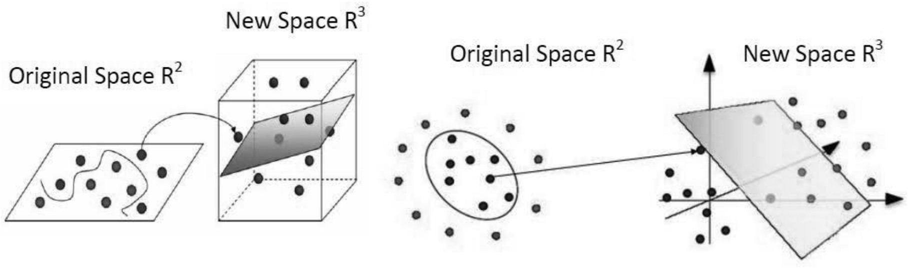


Figure 4.1 Examples of kernel trick mapping

SVMs are a powerful tool for binary classification, capable of generating very fast classifier functions following a training period. There are several approaches to adopting SVMs to classification problems with more than 2 classes (Rifkin & Klautau, 2004). The most common approaches are those that combine several binary classifiers and use a voting technique to make the final classification decision. A more complex approach is one that attempts to build one support vector machine that separates all classes at the same time (Habib, 2008).

One-Against-All (Rifkin & Klautau, 2004) is the earliest and simplest multi-class SVM, as well as the approach adopted by the framework employed in this research. SVM-Multiclass. For a K-class problem, it constructs K binary SVMs. The *i*-th SVM is trained with all the samples from the *i*-th class against all the samples from the other classes. To classify x, the point is evaluated by all of the K SVMs and the class with the largest value of the decision function is selected. Such an approach assures that every point is assigned to a certain class.

4.3 Language technologies and the semantics of data

The previous section described the statistical learning theory and analyzed an efficient algorithm, i.e., the support vector machine. This algorithm learns from annotated data that is represented in a vector space. Because in many phases of the system life cycle, the documents are in natural language, i.e. technical documents, it is necessary to interpret a text to transform it into a vector. Representing the meanings of words and sentences in a form suitable for use by a computer is a central problem in Computational Linguistics (Jurafsky, 2000). The problem is of theoretical interest – to linguists, philosophers and computer scientists – but also has practical implications. Finding a suitable meaning representation can greatly improve the effectiveness of a Natural Language Processing (NLP) system. There have been two distinct approaches to the representation of meaning in NLP. The first, the *symbolic* approach, follows the tradition of Montague in using logic to express the meanings of sentences (Dowty, Wall, & and Peters, 1981). The logical representation of a sentence is built up compositionally by combining the meanings of its constituent parts. In contrast, the *distributional* approach uses statistics about the contexts in which a word is found, extracted from a large text corpus, to provide a vector-based representation of the meaning of an individual word.

The term *distributional* can be interpreted as *context-theoretic* since it encloses a rich family of approaches to semantics that share a usage-based perspective on meaning, i.e., they assume statistical distribution of terms in context plays a key role in characterizing their semantic behavior. Distributional models rest on the hypothesis that the degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear. Therefore, the more the linguistic context is representative of the distributional behavior of a term, the more semantic properties can outcrop inspecting it. Thus, distributional models are inherently related to statistics theory as well as to vector algebra.

With regard to systems engineering, and specifically to requirement analysis, one could imagine that in an evaluative text, term distribution may affect the sentiment expressed. Thus, an interesting question would be if vector based models, representing statistical characteristics of texts, are equally capable as their human counterparts to catch the requirement semantics.

4.3.1 Salton's Vector Space Model

The **Vector Space Model** (**VSM**), developed for the SMART information retrieval system (Buckley, 1993; Manning, 2008) by Gerard Salton (1975), relies on the idea to represent each document as a collection of points in a space, i.e., a vector where each dimension is associated with one term from a dictionary. Points that are close in this space are also semantically similar while points that are far apart are not semantically equivalent. A document is thus seen as a "bag", a set in which repeated elements are allowed, so that not only the presence of a word but also its frequency is taken into account. Hence, the word order is ignored and syntactic information is lost. Vectors are the most useful numerical data structure suitable for contextual representations; geometrical models are employed to capture distributional properties. It is possible to formalize the Vector Space Model as follows:

$V = \{v_0, \dots, v_m\}$ A vocabulary V of all the possible terms is defined. Then m defines the space dimension.

$D = < d_0, \dots, d_m >$ A document D is a m-dimensional vector. Each component d_i is a weight that suggests the presence of the term i in D

Further, VSMs have several important properties. With regard to semantics, knowledge extraction can be automatically performed from a given corpus, requiring much less effort than other approaches, such as hand-coded knowledge bases, ontologies and linguistic based techniques. VSMs perform well on tasks that involve measuring the similarity of meaning between words, phrases, and documents. The document, in our case a requirement statement, is mapped into a space of dimensionality m, that is, the size of the dictionary (usually, a very large number). The contribution of each term in the document is quantified by a function of its presence in the text and over the entire collection of documents, i.e. term frequency (*tf*).

A document-by-term matrix D is defined whose rows are indexed by the documents of the corpus and whose columns are indexed by the terms. The $d_{i,j}$ score gives the frequency of term in the document. The term-by-document is the transpose of D, with the term-by-term matrix given by $D^T D$ while the document-by-document matrix is given by DD^T. This dual representation corresponds to a document view of the problem, while the primal description provides a term view. In the same way that a document can be seen as the tally of the terms that appear in it, a term can be seen as the tally of the documents in which it appears. Text represented by this model often leads to sparse vectors with very high dimensionality. Measuring distances directly under such a representation may not be reliable since it is a known fact that in very high dimensions, distance between any two points starts to look the same. One solution consists of reducing data dimensionality by SVD (see below) and then measure the distances in the new space, which also provides a meaningful representation of semantic concepts.

4.3.2 Word Space Model

The term Word Space is attributed to Heinrich Schütze, who described the model as follows: “Vector similarity is the only information present in Word Space: semantically related words are close, unrelated words are distant” (Sahlgren, 2006: 17) The Word Space model is, as the name suggests, a spatial representation of words. The key idea is that semantic similarity can be represented as proximity in T-dimensional space, where T is an integer number varying from one to very large values. Thus spatial proximity between words indicates how similar their meanings are. In the Word Space model, the similarities between words are automatically extracted from language data by looking at empirical evidence of real language use, i.e. their behavior and relations (Sahlgren, 2006). Therefore, words relations can be classified by their functional differences into syntagmatic and paradigmatic.

Any sentence is a syntagmatic chain composed of significance carrier units. Syntagmatic relationships deal with positioning and co-occurring of entities in the text. In other words, a sentence is a linear relationship consisting of a sequential combination of linguistic entities, or in praesentia relationships. In the first example in Figure 4.2, the verb Read is in a syntagmatic relation with Book, since they are linearly associated to form a correct sentence. In this sense they co-occur. Pages and Read are in a syntagmatic relation as well.

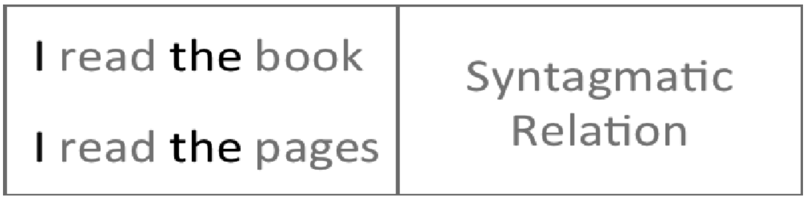


Figure 4.2 Examples of Syntagmatic Relations

On the other hand, a *paradigmatic* relation is an in *absentia* relation, as it relates to entities that do not co-occur in the text. These connections hold between entities that may occur in the same context but not at the same time, and in this sense, items have an equivalent worthiness that makes it possible an in-context substitution. Thus, the words Book, Volume and Article in the second example Figure 4.3 are in a paradigmatic relation, since they can all be substituted in the sentence without affecting the grammar correctness, but one at a time.

Figure 4.3 Paradigmatic Relation

Context vectors from a syntagmatic perspective consist of *n* context regions c_i in the data, with *n* being the dimension of the Word Space:

$$v = \langle c_1, c_2, \dots, c_n \rangle$$

The context region size is the only variable characterizing context vector, and it can appear in either a very small sequence of words or in an entire text region. Obviously, the smaller the context regions are the poorer the statistical foundation will be, leading to a sparse-data problem for the resulting Word Space. Paradigmatic relations may be defined in terms of words that share some of the preceding or succeeding neighbours of a focus word. More precisely three parameters are taken into account:

- The size of the context region observed
- The position of the words within the region
- The context direction with respect to the focus word, i.e. left or right context, Figure 4.4

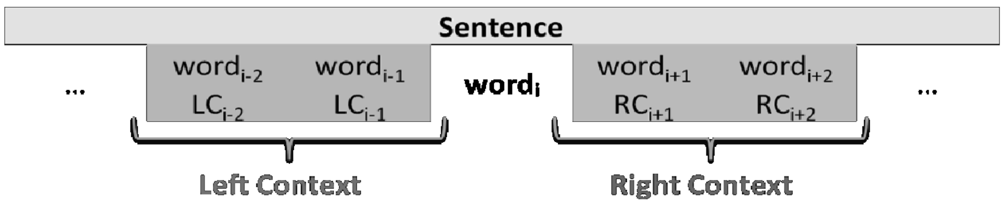


Figure 4.4 Word Paradigmatic Contexts

The context vector is then structured as follows:

$$v = \langle w_1, \dots, w_n \rangle$$

where w_i are neighbor's words. For any focus word, a window is defined, within which information is collected. Schütze and Pedersen (1995,1997) referred to such paradigmatic use of context as *lexical co-occurrences*.

4.3.3 Latent Semantic Analysis

Distributional models can suffer from the disadvantageous problems regarding their dimensions and the sparseness. Such aspects are directly linked to large dataset, where dictionaries, and thus space dimensionality, is significant, but also where any sentence may contain a low percentage of all terms. Many mathematical techniques are used in distributional semantics. **Latent Semantic Analysis** (LSA) is one of the most popular models, based on linear algebra technique for the Principal Component Analysis (Deerwester, 1990; Steinberger & Ježek, 2004), which supports extraction of the most informative dimensions in a matrix of data, thereby solving the problems.

As one of LSA inventors stated:

> *We would like a representation in which a set of terms, which by itself is incomplete and unreliable evidence of the relevance of a given document, is replaced by some other set of entities which are more reliable indicants. We take advantage of the implicit higher-order (or latent) structure in the association of terms and documents to reveal such relationships (Deerwester, 1990: 4)*

Any matrix can be decomposed into factors making use of a linear-algebraic process known as **Singular Value Decomposition** (SVD). The input matrix M is composed of vectors of terms representing evaluative texts, drawing the respective interrelations. SVD allows for less than full ranking of the matrices to approximate the original matrix. LSA needs a matrix of underlining interrelations just mentioned, and this does not require a perfect reconstruction of the original matrix. So, reducing rank to *k*, a smaller matrix can be reconstructed. SVD can be seen as a method for rotating the axes of the r-dimensional space such that the first axis runs along the direction of largest variation among documents; the second dimension runs along the direction with the second largest variation and so forth. The diagonal matrix contains the singular values in descending order. The singular value indicates the amount of variation along the axis. The resulting matrix is the best square approximation of original by a matrix of rank k.

Ideally, the *k* value should be large enough to catch latent features in the data, but small enough to avoid their over-representation. Clearly, reduction in k can remove much of the "noise". However, keeping too few dimensions may lose important information. As discussed in Deerwester (1990), using a test database of medical abstracts, LSA performance can improve considerably after 10 or 20 dimensions, peaks between 70 and 100 dimensions, and then begins to diminish slowly. Thus, in this research, a similar pattern of performance can be expected, i.e., initial large increase and eventual slow decrease, as was observed on many datasets.

LSA supports analysis in a space of lesser dimensions but more semantically informative. In other words, original individual terms are mapped onto a concept space. The sparseness problem for the previous distributional models is solved, and thanks to the SVD, the space dimensionality can be reduced to the most informative dimensions. With respect to the target task, this means that evaluative texts can be represented in a space richer in knowledge, where new relations between terms can be discovered. Such relations add a further generalization power in a classification task. In other words, even a term not seen in the training data could be correctly classified on the basis that a term close to it in the LSA space is a known example.

In this work we apply a short co-occurrence windows of size [−3,+3], centered on the items. This allows for better capturing syntactic properties of words. Lexical items occurring in the collection more than 20 times are considered. The entries of M are the point-wise mutual information between them. The SVD reduction is then applied to M, with a dimensionality cut of $l = 250$. Words are projected into the geometrical space derived through the distributional analysis.. Each word is represented as a vector and then linearly combined in order to have a synthetic representation of the entire sentence.

4.4 Mapping requirements and domain ontologies

One of the research areas involves the feasibility of a semantically oriented approach aiming to automatize relevant inferences involved in the requirement analysis (RA) stage of a system development lifecycle. In order to reproduce results achieved by the analysts and experts, an ontological organization of the domain knowledge is required, to express all products and processes taken into account in the RA stage.

As discussed by Thomas Gruber (1993), an ontology is an explicit specification of a conceptualization. From a theoretical and philosophical perspective, an ontology is a systematic account of existence. For AI systems, what "exists" is that which can be represented. When the knowledge of a domain, such as the subsea domain, is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, the ontology of a program is defined by a set of representational terms. In such an ontology, definitions associate the names of entities, e.g. products (`XMT`) or processes (`Testing_process`), in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms.

It is assumed here that an underlying model is available and it is expressed through a more or less complex domain ontology. Given this assumption, the research verifies the benefits that an ontological organization of the domain knowledge provides to text management (or other unstructured information) during the RA stage. These benefits arise from the direct availability of semantic notions within the texts (e.g. design specification themselves) that support several useful inferences: automatic tracing of requirements insisting on the same components; alignment of requirements acting on similar physical properties; clustering of similar, but possibly inconsistent, requirements spread across large documents. In this research, broad use is made of natural language processing and machine learning techniques to automatize semantic annotation of requirement documents by locating sentences expressing requirements and assigning them specific (ontological) types.

4.5 Domain analysis

This section reports the results obtained during the first part of the project (namely the domain analysis phase). These activities are the foundation for the ontological model (OM) design.

The aim of the research is to improve project execution by providing an AI assisted requirements extraction environment as addressed by Basili, Moschitti, Giannone et al. (2005,2007,2008,2011) , improving the full compilation of the requirements register with the required quality and within the tight time constraints imposed by the short cycles typical of the subsea oil industry. The project documentation is intricate due to its depth and fine-grain nature with respect to the contents that are highly technical and focus on complex engineering knowledge, as well as for its sheer size that make the process of locating relevant information and navigating across multiple and heterogeneous documents challenging.

Semantic technologies are an emerging area of research in the context of system engineering. In recent years literature shows an increasing interest into the application of semantic technologies in the design support during various phases of the life cycle of a system. For example, the requirement analysis is a relevant application area for a variety of semantic technologies related to the extraction, disambiguation and exploitation of the knowledge derived from technical requirement documents. Most methods rely on shallow-language processing technologies for the automatic extraction of core concepts (e.g. components/devices, their parts and functionalities) and norms (e.g. constraints on the use of components).

Uschold and Gruninger (2004) demonstrated that machine-readable models of the domain semantics increase the ability of the system to reason about their data. Machine learning allows automatizing the detection of most of the inferences involved in an application task. This research uses the application of state-of-the-art learning algorithms (e.g. SVM) as a cost-effective approach to classify and validate requirements from natural language text in support of requirement analysis for the design of a complex systems and illustrates the application of learning methods to this scenario. Early results show the robustness of these methods to the targeted scenario and suggest their applicability to other domains or scenarios that are interesting for NSC and other companies.

This research is the first application of semantic technologies for the (semi)-automation of the requirement collation and analysis phase within the subsea oil extraction project domain. Its aim is to handle requirement and tender documents and automatically extract the information needed during the requirement analysis phase. This required the architectural design of a solution that processes unstructured documents, i.e. written in natural language, and enriches texts with linguistic information with rich semantic content. Through this approach, those sentences expressing one or more requirements are retrieved and the target information is extracted and presented to the analyst in an understandable and easily accessible way.

Given the assumption that an ontological organization of the domain knowledge exists, in terms of relevant processes and products, it is possible to map requirements (expressed in form of sentences or other unstructured information) to existing concepts. As a first example, let us consider a sentence expressing a requirement such as "*Self tests of the XMT power system shall be implemented*". Given an ontology describing all products and processes involved in the oil extraction process within the subsea domain, it is possible to map such sentence to a concept representing the `christmas_tree` or a process such as `testing`. As the ontological model is designed to represent all properties and functionalities of the products/processes, the resulting *reasoning* system can suggest to the analyst all product subparts that are involved in the test, or other requirements that have been previously activated and are related to the testing phase. These requirements are used to populate the *requirement repository* that can be later easily accessed by the analyst. The adoption of a strong modelling of the subsea domain will enable the automatic analysis of the extracted information to recognize/acquire existing dependencies or inconsistencies among the repository requirements. The study analysed the implications of using an AI approach by examining the results obtained through the study against a recently fulfilled project, assessing the differences in requirements elicitation and early implementation. The evaluation was carried out by extracting the requirements through semantic technologies and assessing the results through review by a panel of long time senior domain practitioners whose assessment delivered the performance benchmark. As a prerequisite, the work developed an ontology forming the basis of the semantic analysis. The ontology was validated by industry domain experts.

One important objective of the current research is the definition of an ontological model to support the enrichment, mining of the documents produced and used by NSC, from tender texts to catalogues, using the methods previously described, the aim was to provide rich explanatory descriptors of targeted phenomena (instances of requirements) in order for them to be located automatically within the documents. The proposed subsea ontology caters to a number of general concept `type`(s) able to describe classes of entities in the real world (such as `physical phenomena` or `physical entities`, e.g. the *pressure* vs. a *tube*, or `abstractions` vs. `activities`, e.g. *numbers, percentages* vs. *filling* or *moving actions*). General types are then used to annotate texts and sentences within the documentation by semantic information, i.e. concepts, relations and properties useful to characterize design choices (as in specification documents), requirements (design documents) or product descriptions (catalogues). The automatic recognition of such semantic phenomena enables more expressive search operations, and comparisons, across large amount of documents. The main objective it to elicit requirements (found in general sentences) such as "*the pump XYZ is designed to resist seizing at low temperatures* (*-40° F.*)" or "*the minimum working temperature of the pump is -40°*" that can be translated into a machine-readable form, such as `design_temperature(pump(XYZ),-40°K)`. Then, as an example, the ontological enrichment of texts according to the proposed schema would enable semantic queries, such as "*report all temperature constraints for xmas trees*".

The advantages that an ontological organization of the domain knowledge provides to NSC analysts are several: direct availability and identification of semantic notions within the texts (e.g. design specification themselves); support to several useful inferences, such as a concept-driven search within the requirement register; and, alignment and versioning of specifications, requirements and checklists during the analysis stages.

The OM represents the set of targeted information the proposed solution is expected to find in the main contractor documentations, and use to automate the requirement analysis process. Specifically, the domain analysis is the basis of three major threads in the project: design basis study, ontology design and population, and development of the versioning system.

Having constructed the ontological model, the system is able to acquire main contractor documentation, extract the useful context and apply NLP as well as machine learning techniques in order to assign specific *semantic types* to texts reflecting targeted information in the RA activities within subsea domain. These types are defined in the OM and they can be used as semantic meta-data to help retrieve the requirements, thereby enabling a semantic search of texts expressing specification of requirements about products or functionalities.

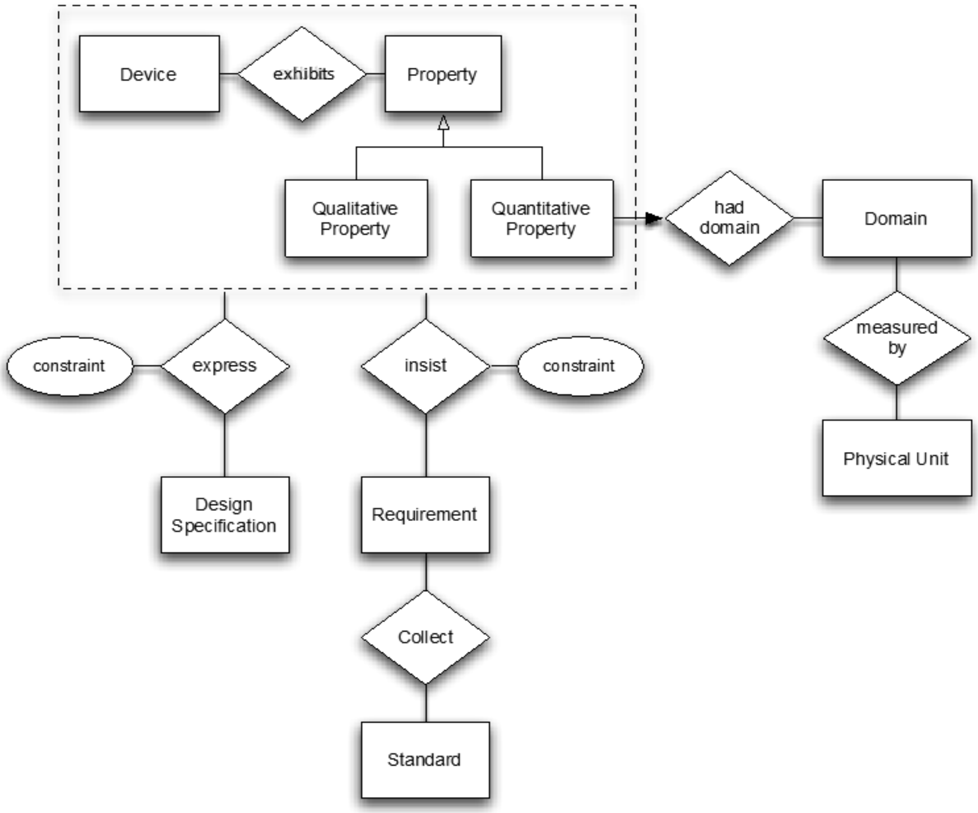


Figure 4.5 An Entity Relation graph representing the requirement ontological model

4.6 An ontological model of requirements

The set of targeted concepts in this project is derived from the following abstraction of requirements that are formalized in the entity relation diagram shown in Figure 4.5. Requirement properties are analysed from different perspectives, defining the following semantic types to be assigned to each requirement.

Requirement. According to our model, a `requirement` *insists* on a `device` that *exhibits* some `property`. A `device` is an object, a machine, or piece of equipment (such as an *XMT* or a *manifold*) that has been designed and manufactured for some specific purpose. A property is any characteristic of a device that is observable and objectively verifiable.

Requirements are collected in standards and each property that is exhibited by a device refers to a specific design specification; as an example a requirement such as "The XMT should operate at +80° Celsius" refers to the `design_specification` called "`design_temperature`".

Standards. When a specific standard, e.g. from the ISO or ASTM is mentioned within a document, it is referenced to the specific `Standard` class and is assigned as meta-data.

Relevant standards published by the following internationally recognized organizations have been considered: ISO, ASTM, NACE, DNV,

As an example, a standard is expressed in bold in the following sentence "This part of **ISO 13628** provides guidelines for the design".

Devices. Each requirement is expressed with respect to one or more devices. The NSC provided a taxonomy defining the entire set of existing devices covered by the supply chain, called MACS. The complete taxonomy is reported in the Appendix 1 – The MACS classes.

Property. a property is a characteristic of an object. A physical property is measurable, its value describe a state of a system and is often referred to as observable.

- `Quantitative_Property`: a property that can be directly expressed and measured. The measure can be expressed through a specific numeric value. Each quantitative property is expressed with respect to a Domain e.g. "temperature" and is thus measured with respect to a specific unit, e.g. "40° Celsius".
- `Qualitative_Property` (**-ilities**): a property that is observed and cannot be generally or easily measured with a numerical result, e.g. durability or interchangeability.

Domain. In this work the term domain is to be intended as the "domain of discourse", i.e. is the set of entities over which certain variables of interest in some formal treatment may range

Unit. When a requirement refers to a physical property of the considered device, a specific physical quantity must be explicitly mentioned as a unit with a specific measure applying quantitative constraints to the requirement. Twenty-one physical units of measure, deemed the most relevant for the domain, are automatically retrieved in order to associate the `physical_quantity` semantic type. As an example, a measure of length is expressed in bold in the following sentence "*Pipe length shall be **10 m (32,8 ft)** or more.*" The complete set of physical quantities applied in this work is reported in the Appendix 2.

Design specification. When an analyst is required to process the requirements expressed in a document, such as a tender, s/he is also required to group the requirements in a set of classes, here called design specifications, that characterize the set of requirement that should be considered in each document. As an example, given a document an analyst should consider all sentences expressing requirements related to the `design_pressure` or `safety_system_requirements` classes. Being able to associate each sentence to one of the above classes, a semantic search system that filters the document to be analysed can easily be implemented, largely reducing the collation effort required by an analyst. While the automatic association between texts and such design specification classes has been modelled as a classification task, the complete set of design specifications considered in this project is reported in Section 5.2.

4.7 Architectural design

Reveal srl proposed an integrated solution that is designed to acquire documentation required in the requirement analysis phase, extract the useful context and apply NLP and Machine Learning techniques in order to assign specific ***semantic types*** to the extracted requirements in order to associate them to the ontology of the subsea domain. These types can be also used as semantic meta-data to help the retrieve of the requirements, so enabling a semantic search of texts expressing specification of requirements about products or functionalities.

In Figure 4.6 the architecture is shown in order to emphasize the different modules involved. The dotted line boxes delineate interacting modules.

- The **import handler** loads and pre-processes the requirement documents, in order to acquire a representation that is readable by the following modules.
- The **versioning handler** manages the documentation versioning. When the upgrade of a given document is available, it retrieves differences between the documents in order to enable the tracking of the requirements across upgrades. As an example, when a sentence expressing a requirement is added, the system will assign a proper type to the new sentence, and it will thus notify the addition of a novel requirement.

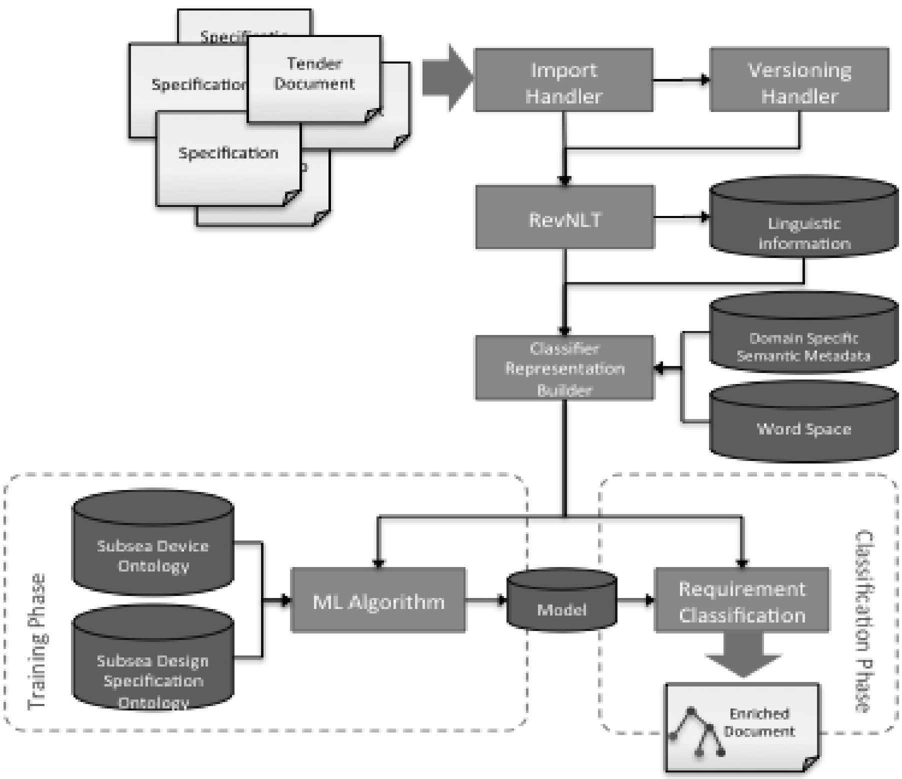


Figure 4.6: System Architecture

- The **reveal natural language toolkit (RevNLT)** implements techniques of natural language processing (NLP) in order to achieve a morphosyntactic analysis of the nature of the texts contained within documents. An example of this analysis is the segmentation of the documents into sentences, the identification of the main classes that characterize grammatically the words that make up the sentences (e.g. nouns, verbs or adjectives); and, parsing of sentences, which allows the extraction of linguistic constructs such as the *subject-verb-object complement*. In the overall architecture, this system represents a module providing linguistic information useful to build artificial representation for the learning algorithm.
- The **classifier representation builder** acquires each sentence, and selects from the NLP system the proper information needed to generate the artificial representation useful for the learning algorithm, as discussed in Section 4.3. Some information can be derived by lexicon and terminology collection that represents *domain specific semantic metadata*. The lexical generalization is acquired through the *WordSpace* model introduced in Section 4.3.2.

Each representation follows a different workflow depending on the training or the classification phase. When acquiring the statistical model data is expected to be labelled with *types* derived by the ontologies, i.e. **subsea process ontologies** or **subsea product ontologies**. The **machine learning algorithm** loads labelled material in order to acquire a function to associate novel requirements and proper semantic types. This statistical ***model*** is then used by the **requirements classifier** to associate proper semantic type to novel sentences.

The Reveal proprietary RevNLT has been customized to consider the following (domain-specific) information:

- over 700 MACS classes, such as the class "*7010300 - subsea XMT sub-Assemblies.*"
- over 2,000 lexical variants of the involved devices. As an example "sub-assy subsea xmt sub-assembly", "christmas tree", "xmas tree", "xmt", "xm tree", "xt","sub-assy xmt", "sub-assy xm tree", "sub-assy christmas tree", "christmas trees", "xmas trees", "xm trees", "sub-assy xm trees", "sub-assy christmas trees" are lexical variations of the devices represented in the above 7010300 class
- Specific **automata** to cover unlisted EN and ISO standards
- Specific **automata** to recognize the involved 20 physical quantity classes considering more than 800 lexical variations of the above physical quantities.
- A word space model, built according to the process described in Section 4.3.2 that contains a geometrical representation for more than 20,000 lexical items.

5. Case studies

The purpose and value of a proper and accurate requirement extraction, classification and allocation to the correct entity is discussed in the papers “Subsea oil production: The design basis” and “The case for requirement management: How industrial design requirements are specified and executed in the Oil Subsea Production Systems (SPS) vs. Shipbuilding industries” which are part of this work.

To substantiate the suitability of ML assisted extraction methods results for industrial applications quantitative measures are necessary, including comparison with suitable benchmarks. This chapter is dedicated to providing appropriate measurements for evaluating the research approach effectiveness in the chosen industrial domain. The proposed methodology is applied to a real industrial scenario and documentation derived from within the subsea domain is utilised. The aim is to show the applicability of NLP and ML techniques in characterizing requirements and assigning them proper types that reflect the devices identified and the design specifications involved in the specific requirements; these classes are expected to populate the subsea domain ontology. In order to provide an exhaustive and quantitative evaluation, the NSC has provided a dataset of requirements, labelled according to the targeted classes.

In Section 5.1 the achievable quality of the sematic annotation discussed in Section 4 is reported on real data sourced from main contractor documentations. In Section 5.2 the semantic search engine augmented with the semantic annotation is shown and discussed. Finally, in section 5.3 the first results of the document versioning module are presented.

5.1. Case Study 1: Modelling requirements with kernel based learning

An empirical evaluation has been carried out to determine the potential quality of the adopted classification paradigm, i.e. the quality in the association between texts and concepts representing design specifications. The classification module uses a support vector machine classifier (see section 4.2) to determine the design specification class expressed in a sentence. Different approaches are adopted to derive a computational model of the linguistic information expressed in sentences to investigate different information suitable for the target task and identify the most informative representation for the learning algorithm. Let us consider the requirement “*self tests of the xmt shall be implemented*”: four different computational representations for the SVM learning algorithms can be adopted

- The **full-text *bag-of-word*** (ft-*BoW*) model: counts mainly on lexical information, resulting in a vector whose dimension reflects the different words composing each requirement. As an example, dimensions are “*self*” “*test*” or “*XMT*”;
- The **linguistic processed *bag-of-word*** (lp-BoW) model: each sentence is processed by the NLP system; it allows providing more expressive information to the learning algorithm, exploiting external and domain specific semantic information. As an example, the dimension “*self test*”, carries out a more specific meaning of “*test*”; or “*christmas tree*” that is the meaning of the acronym XMT. This will extend the simpler information provided by the ft-*BoW*. The linguistic analysis also normalizes information, avoiding data sparseness. As an example the verb “*implemented*” is replaced by the infinitive version “*implement*”;

- The **entity-driven *bag-of-word*** (ed-BoW) model: expresses the linguistic analysis filtered by the word part-of-speech. It focuses more on the objects or devices, rather than filtering out only nouns;
- The **Latent Semantic Analysis** (*LSA*) model: counts on lexical information enriched with expansion distributional models to represent meaning of words. The word space is derived through the distributional analysis of a document collection made available by the NSC as well as from a page collection downloaded from relevant publicly available sources.
- Each representation enables the application of a different kernel functions. The first BoW models are used in linear kernels, while the LSA is used in a gaussian kernel. The combination of kernel functions is applied: consequently, the kernel combination

$$\alpha\,(ft\text{-}BoW) + \beta\,(lp\text{-}BoW) + \gamma\,(ed\text{-}BoW) + \delta\,(\text{LSA})$$

linearly combines lexical properties captured by BOW kernel and generalized by LSA. Here, parameters α, β, γ and δ weight the combination of the three kernels.

5.1.1. Evaluation Metrics

Performance evaluation in automatic classification system is needed to establish if the model acquired from the examples is suitable for use on future cases (usually called test cases) so that in the operational scenario the system is undistinguishable from the ground truth and the classification function has been perfectly learned. The observation of the outcomes of a classifier $h(x)$ allow to compute the following scores:

- *Error rate*, that is the ratio between the number of test cases not correctly classified (whereas the system prediction is not in agreement with the gold standard) and the total number of test examples processed. its complement,, the percentage of the correctly classified examples constitutes the so-called *accuracy*;
- *Precision* and *recall*, that better account for the system tendency to introduce mistakes (false positives) or miss correct decisions (i.e. false negatives), depend on a category C_i (for example, the *i*-th emitter E_i in the library), and the following quantities[1]:
- the number *tp* of *true positives*, i.e. correctly classified examples, that is the number of examples $x \in T$ such that $C_i \in h(x)$ and $C_i \in GS(x)$;
- the number *fp* of *false positives,* that is the number of examples such that $x \in T$ and also $C_i \in h(x)$ with $C_i \notin GS(x)$;
- The number *fn* of *false negatives*, that is the number of examples such that $x \in T$ with $C_i \notin h(x)$ and $C_i \in GS(x)$.

[1] GS represents here the *Gold Standard,* that is the collection of annotated examples *x* labeled in the different classes C_i (when $C_i \in GS(x)$) or outside these classes (i.e. $C_i \notin GS(x)$).

Given the above definitions also the following definitions hold:

$$Precision = \frac{tp}{tp + fp} \qquad Recall = \frac{tp}{tp + fn}$$

In general, *Precision* and *Recall* vary in an inversely proportional manner. Moreover, as a single measure is often preferable, a largely adopted combination function is the harmonic mean of precision and recall that gives rise to the following score called F1 measure:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

5.1.2. Experimental Results

The first use case presents the evaluation of the acquired classifier with respect to the set of specific design specification classes defined within this project. The aim is to acquire a classifier, i.e. a function, able to associate each sentence expressing a requirement to a specific design specification class.

The SVM learning algorithm has been employed. As the association with targeted classes is a multi-classification task, several binary SVM are trained according to the one-vs-all schema (Rifkin & Klautau, 2004). Each novel requirement is processed by all binary classifiers, each one representing a possible class. It means that, when a classifier provides a positive response, the corresponding class is assigned to the text. When no classifier expresses a response, the text is not recognized as a requirement. Labelled sentences from a training corpus provided by the NSC have been considered.

The following two-level taxonomy has been considered to represent design specifications:

- a first conceptualization level of 7 classes has been considered, e.g. `conops` and `controls`.
- each concept of first level first has been specialized according to second level concepts. As an example, the first level concept `CONOPS` is specialized by further conceptualizations, such as `design weight` or `environmental conditions`. When a requirement is associated to a concept of level 2, it inherits also the associated concept of level 1. The total number of second level concepts is 54.

The complete list of considered concepts is reported in Table 5.1

The cooperation with NSC analysts allowed the setting of a benchmark through manual annotation of a design basis describing a christmas tree, namely the document TR3571. The analysts labelled more than 1,000 sentences linking them to one of the above concepts. Even though this is still a reduced set of data, it is a starting point to show the potential contribution of the proposed classification paradigm.

Two different experiments have been carried out to consider the quality of the adopted classification schema:

- a multi-class classifier linking sentences to one or more concepts from the first level of the Design Specification classes.
- a multi-class classifier linking sentences to one or more concepts from the second level of the Design Specification classes.

Table 5.1 The complete set of Design Specification Classes considered in this project

Concept Lev1	Concept Lev2
CONOPS	Design Weight, Environmental Conditions, Installation, Installation Tooling, Loads, Manifold, Overtrawlability , PLET, Production, ROV intervention design criteria, Sand Production , Soil Conditions, Tagging, Testing, Tie-In Tooling, Tubing Hanger, WAG System, Water and Gas Injection, Water Injection, Wellhead PGB and Casing Hanger, XMT
CONTROLS	Electrical distribution system , Hydraulics, Umbilical Function List, Umbilical System
DESIGN CODES	Design Life, Design Pressures , Design Temperatures , Design Water Depth , Dropped objects, International standards , National regulations , RAMS, Safety System Requirements, System Availability, Systems shutdown levels, Technical Risk and Safety Management, Water depth , Wellhead Shut-in Pressure
FIELD LAYOUT	Coordinates, Field layout
FLOW ASSURANCE	Chemical distribution system , Flow forecast , Fluid Composition , Gas Lift/Injection , Insulation and Hydrate management, Methanol distribution
LCI	LCI
MATERIAL SELECTION	CO2 Content, H2S Content, Material Selection, Seawater for Water Injection Specification

When the first level of concept is conserved, Table 5.2 and Figure 5.1 show the number and the distribution of training examples for each class. Classifier parameters are tuned and evaluated with a leave-one-out strategy on the dataset: at each time an example is used as test example (used to estimate the evaluation metrics) and the remaining examples are considered as training set.

Table 5.2 Number of training example for each class when considering first level classes

Class	#Examples
CONOPS	1469
DESIGN CODES	359
CONTROLS	270
FLOW ASSURANCE	107
MATERIAL SELECTION	93
LCI	40
FIELD LAYOUT	16

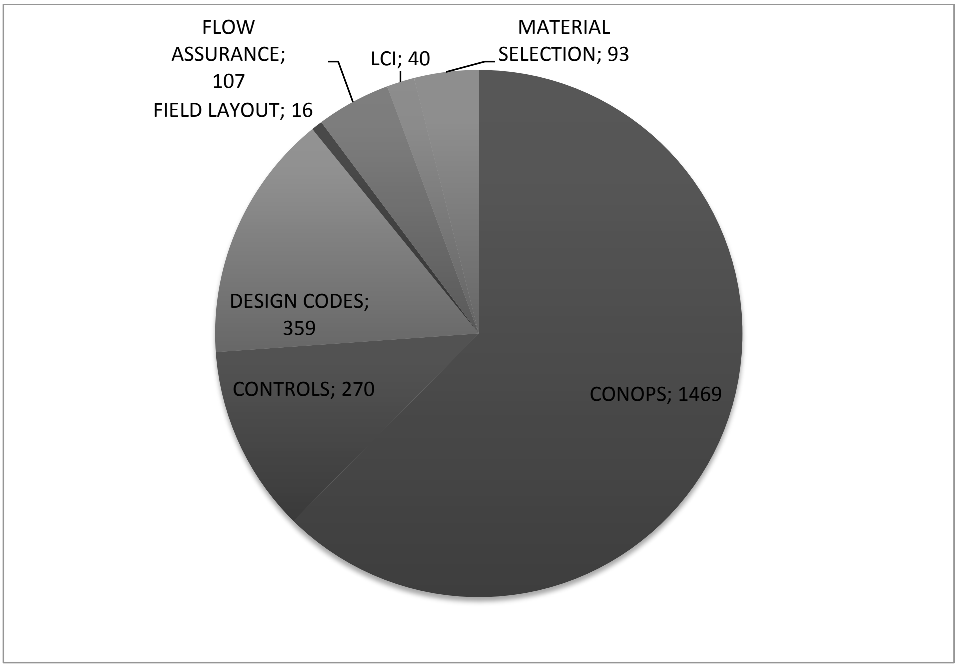


Figure 5.1 Data distribution when considering first level classes

Table 5.3 Statistics and results when considering the first level concepts

Number of requirements	1,024
Number of (requirement, class) pairs	1,229
Number of classes	7
Average Number of classes for sent	1,2
F1 Baseline Random	16,0%
F1 Baseline Most frequent class	52.7%
F1 of the SVM classifier	86.8%
Percentage of sentences labeled at least one correct class	91,7%

When the second level of concept is conserved, Table 5.5 and Figure 5.2 show the number and the distribution of training examples for each class. Again, classifier parameters are tuned and evaluated with a leave-one-out strategy on the dataset.

Table 5.4 shows the overall statistics in terms of number of training example. As each requirement can be labelled with more that one class, the number of single requirements is less than the number of possible (*requirement*,*class*) or (*r*,*c*) pair. Moreover, Table reports also the results in terms of F1. The same table reports also the two baselines, i.e. when classes are assigned randomly and when each sentences has been assigned to the class more represented in the training material, i.e. the `Installation` class. Again, the percentage of sentences labelled at least one correct class is reported.

Table 5.4 Number of training example for each class, considering second level classes

Class	#Examples
Installation	373
Installation Tooling	262
XMT	212
DESIGN CODES	148
CONTROLS	133
loads	130
Tubing Hanger	126
Wellhead PGB and Casing Hanger	103
Material Selection	93
Testing	85
Hydraulics	82
ROV intervention design criteria	65
Chemical distribution system	60
Electrical distribution system	49
Safety System Requirements	49
Design Life	46
Tie-In Tooling	42
Design Pressures	41
LCI	40
Environmental Conditions	36
Methanol distribution	29
Design Temperatures	19
Design Weight	17
RAMS	17
Field layout	16
Design Water Depth	15
Flow forecast	13

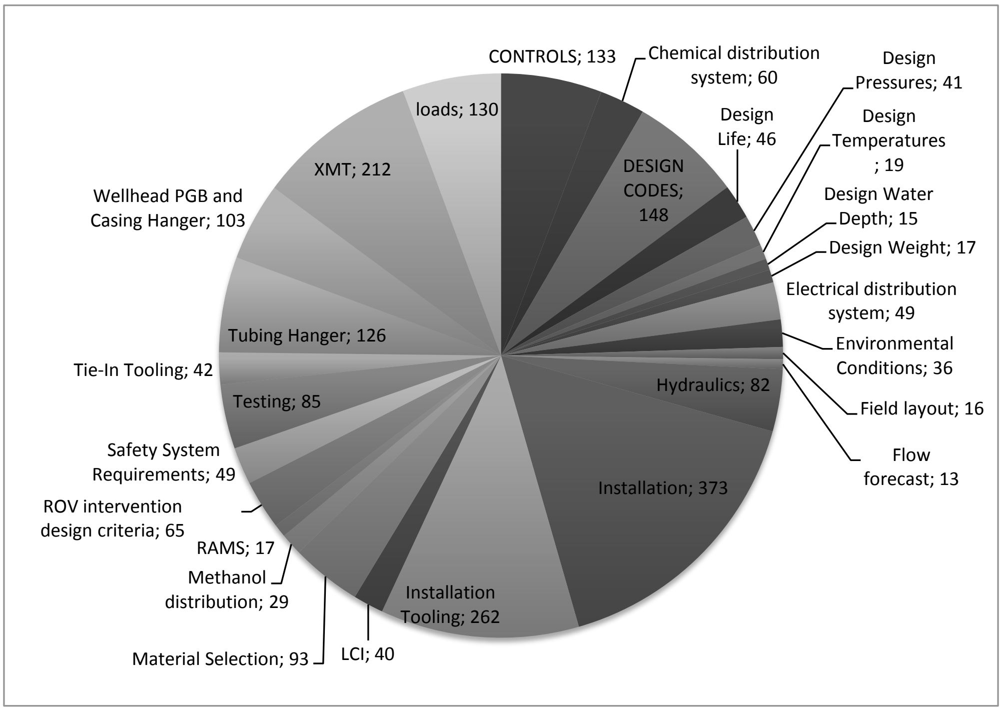


Figure 5.2 Data distribution when considering second level classes

Table 5.5 Statistics and results when considering the second level concepts

Number of requirements	1,024
Number of (requirement, class) pairs	2,301
Number of classes	27
Average Number of classes for sent	2,2
F1 Baseline Random	5,6%
F1 Baseline Most frequent class	22.7%
F1 of the SVM classifier	66.6%
Percentage of sentences obtaining at least one correct class	89.5%

The results show a significant improvement of classification quality when the SVM classifier is adopted. This improvement is less evident when the first level of concept is considered, i.e. 86% vs 52% of F1. When the second level of concept is considered, the improvement is more noticeable, i.e. 66% vs 22%, as a higher number of classes makes this a very challenging task. This task is even more challenging if considering that each sentence can be labelled with a combinatorial number of classes, as a sentence can be labelled with 1, 2, 3 up to 54 classes. That the overall result of 66% obtained through the second level classes without deeper levels support is very promising, particularly considering that only 1,000 examples have been labelled for a set of 54 possible concepts, with an average of less than 18 examples for each class. This result is significant, when considering that about 90% of sentences have been labelled at least with a correct class.

5.2. Case study 2: the semantic search engine

A semantic search engine has been implemented customizing the Reveal search engine to exploit the ontological information derived by the semantic processing carried out in this project in order to:

- Retrieve documents (as well as passages or sentences) from the NSC document collection that refer to specific concepts.
- Enable a semantic filter of the retrieve material (e.g. selecting only sentences related to the analysis of the “design pressure” of an “XMT”).
- Enable advanced reports from the retrieved material (e.g. clustering results of a search engine by considering the specific requirements involved in the retrieved texts).

The development of the user interface (UI)has been carried out taking into account the following guidelines:

- **Expressivity**: The user interface must expose all information extracted from text:
    - Differences between sentences
    - Discovered entities (standards, components and physical quantities)
- **Navigability**: The UI must allow users to navigate within the results provided by the system being always able to recover the set of concepts discovered by the system.

The system has been recently deployed in its full functional version, so that analysts will be able to quantitatively and qualitatively evaluate the impact of its semantic capabilities, and the overall usability of the personalized search environment in a systematic manner with respect to realistic data sets and main contractor documentation. At the moment of writing the following 17 documents have been semantically analysed within the system.

Table 5.6 List of documents semantically analysed in this research

ISO- 13628_10	ISO 13628_2	Design_Basis_XT_Morvin
ISO 13628_11	ISO 13628_7	TR0038_Version_4
ISO 13628_14	ISO 13628_9	TR2382_Ver_4_Material_and_fabrication_requirements_for_subsea_XT
ISO 13628_15	ANSI_API_SPECIFICATION_6A	TR3101_v3
ISO 13628_16	AO_100_00_KAMG_000005_rev03	TR_1053_ver_8
ISO 13628_17	CG_MHN_23_0001_744119_00	

5.2.1. The Semantic Search front-end

In the Search tab, contextual search and query completion is offered to the user. In Figure 5.3 the semantically driven query expansion is shown. Suggestions related to the keyword "*XMT*" provided by the user are shown, where alternative ways of expressing the query term are proposed, such as "*christmas tree*" or semantically related nouns, such as "*wellhead*" or "*tubing hanger*" are the proper continuation of the query, given the underlying domain, i.e. the subsea domain.

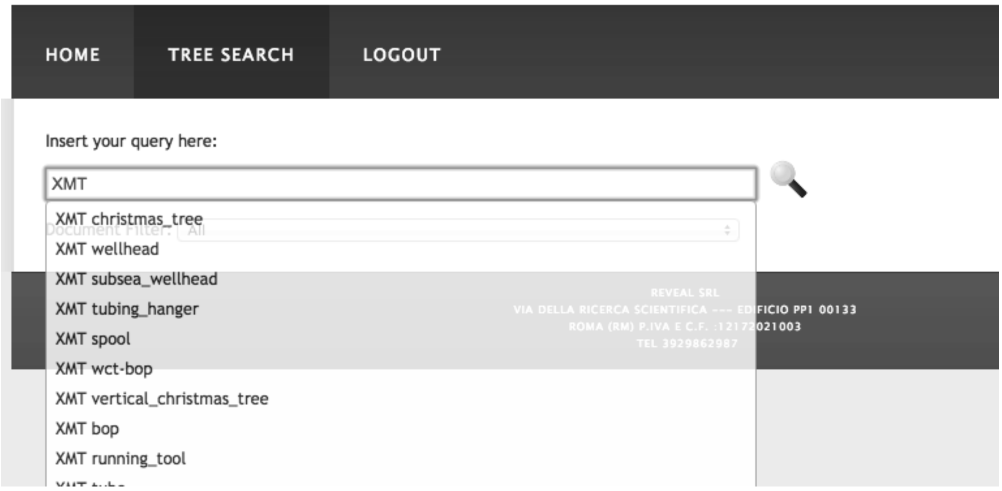


Figure 5.3 Example of Query Expansion

The search activity can be performed by selecting a subset of documents to be considered, in order to focus only to those texts to be analysed, as shown in Figure 5.4.

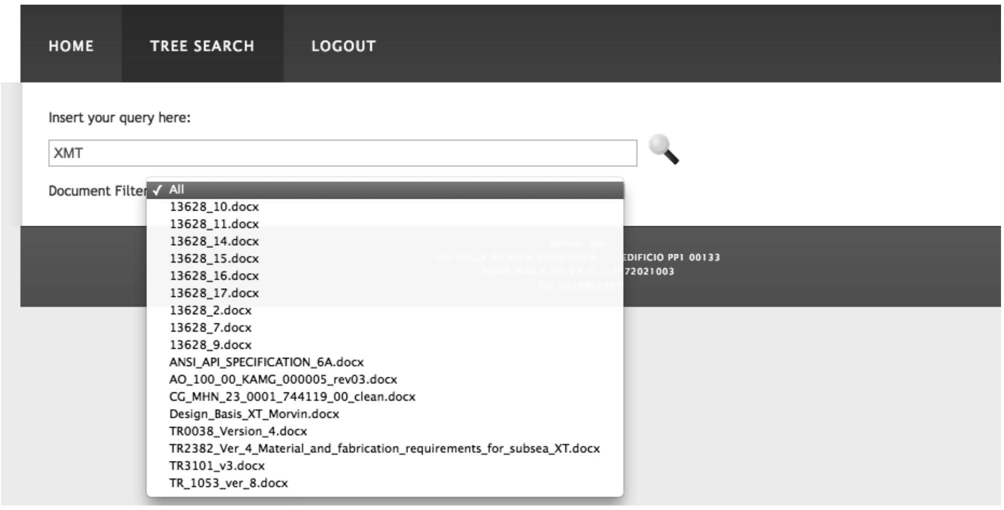


Figure 5.4 Documents selection

When the query is submitted to the system, the set of paragraph that semantically matched to the query are retrieved. Notice the system does not perform a simple string match, as for example, given a query terms like "*XMT*" it allows retrieving also documents that contains alternative ways of expressing the query terms (the lexical variation).

Figure 5.5 screenshot of a subset annotated texts retrieved by the system with the query "fastener"

Figure 5.5 shows an example of annotated texts retrieved by the system with the query "*fastener*", where the documents are annotated with colours that highlight concepts retrieved in documents, such as the physical quantities expressing temperature in red. It is worth observing that the system also shows on the right column an excerpt of concepts recognized in the set of documents retrieved, i.e. a subset of the subsea ontology discussed in Section 4.5. In particular, the subset of concepts expressing devices (derived from the MACS hierarchy) is reported, such as "`subsea XMT sub-assemblies`" or "`skid-transport`". When one of those concepts is selected, the system shows the subset of texts that answers to the query and expresses at the same time the selected concepts. As an example, in Figure 5.6, the first two results retrieved by the system after submitting the query "*XMT*" and selecting the "`pressure`" concept are shown.

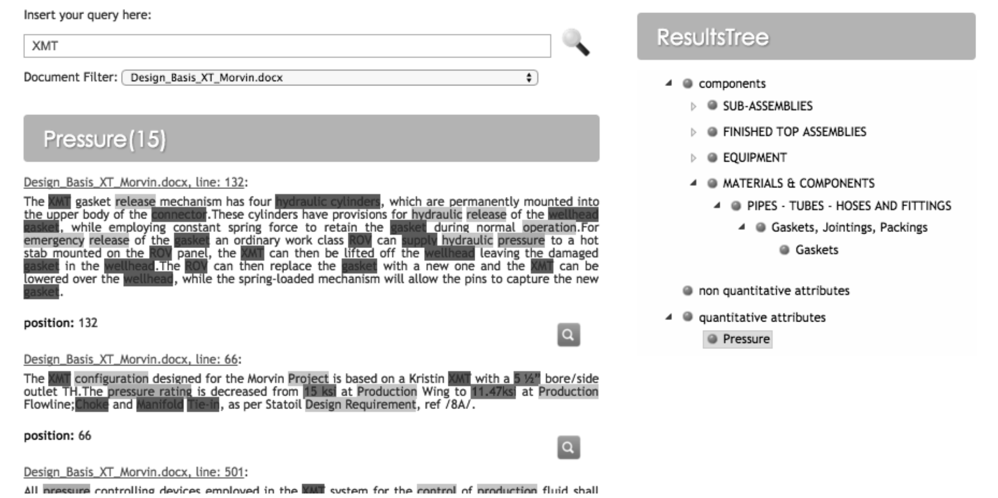


Figure 5.6 Screenshot of a subset System results for the query "XMT" when the "pressure" concept is selected

While the above two examples reports a query term reflecting a device, the semantic search engine also allows other types of queries. Figure 5.7 shows the system responses for the query "*working_pressure*", i.e. a design specification when the device class of "`metal_clips`" is selected: the retrieved snippets show explicit reference to pressure measure, e.g. the physical measurement of "*69 Mpa*" or "*100 psi*", together with the mention to metal clips such as "*rings*". Finally, Figure 5.8 shows the system response for the query "*manifold*" (i.e. a device) when the "`testing`" design specification is selected.

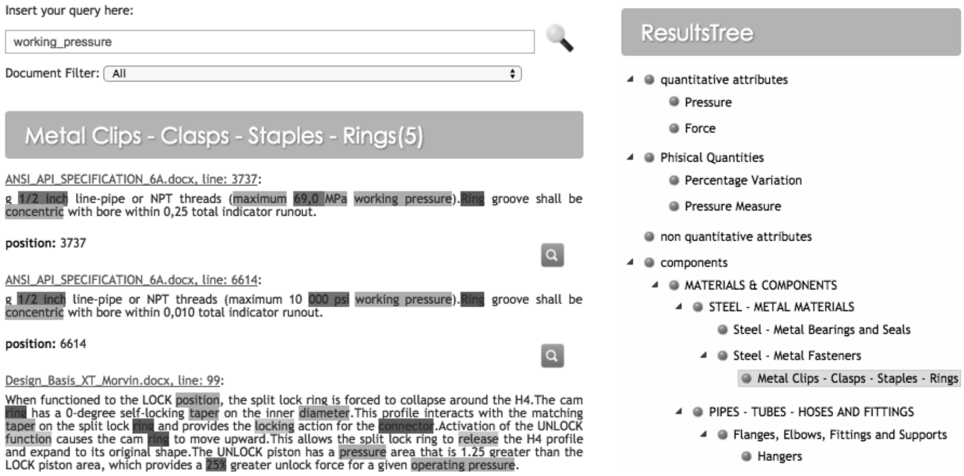


Figure 5.7 screenshot of a subset System response for the query "working pressure" when the device class of "Metal_Clips" is selected

In the last example, the retrieved snippets refer to manifold and express at the same time requirements inherent the testing activity, such as "*function test*" or "*testing*".

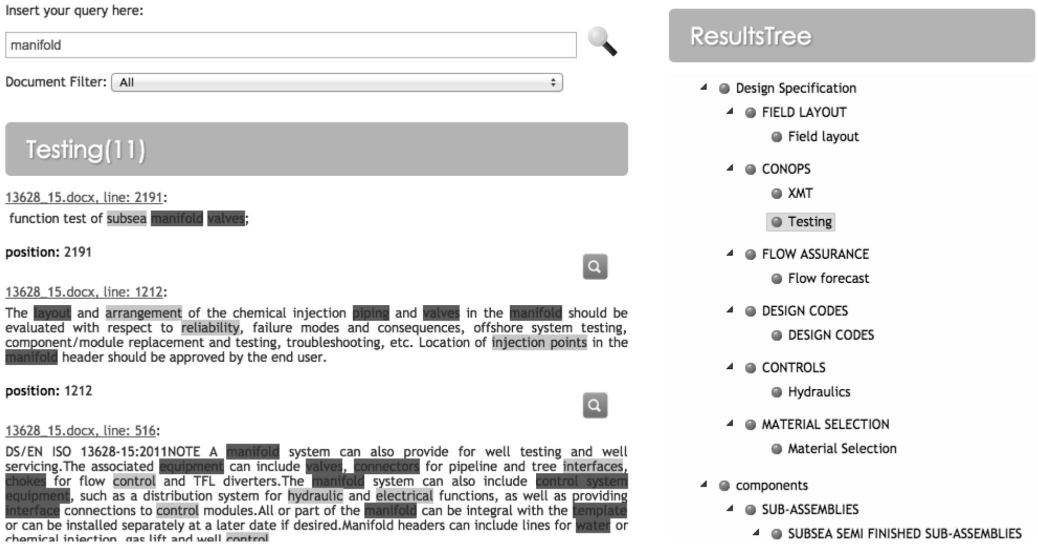


Figure 5.8 Screenshot of a subset System response for the query "manifold" when the "testing" design specification is selected

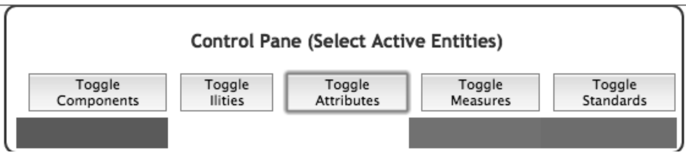
5.2.2. Browsing semantically annotated main contractor documents

From the semantic search front-end, by clicking on each retrieved snippet it is possible to directly access the main contractor documents, e.g. the one shown in Figure 5.9 where, all sentences from the original document are shown and the retrieved, mentions to the devices are marked in green (such as "*XMT*" or "*production master valve*"), the relevant physical quantities are marked in red (such as "*15 ksi*" or "*500m*") and the standards are marked in grey (such as ISO 13628/ 10423).

5.3. Case study 3: document versioning

The application of automatic anchored ***document versioning*** aims at supporting analysts in the trace of requirement change across documentation. When the upgrade of a given document is available, it retrieves differences between the documents in order to enable the tracking of the requirements across upgrades. As an example, when a sentence expressing a requirement is added, the system will assign a proper type to the new sentence, and it will thus notify the addition of a novel requirement.

Table 5.7 shows an example of result of the automatic versioning proposed and investigated in this project. In the first column the number of paragraph in the original document is reported. The second and third columns represent the version *k* and *k+1* of a document. When something in the paragraph changes it is emphasized by the text color, as shown for the paragraph 137. When paragraph is added or removed a blank section is reported, as shown for the paragraph 140.



…

64 Xmas Tree

65 The Subsea Production Xmas Tree shall be of a side valve configuration, as detailed on the XMT Schematic general arrangement, ref /11/.

66 The XMT configuration designed for the Morvin Project is based on a Kristin XMT with a 5 ½” bore/side outlet TH.

67 The pressure rating is decreased from 15 ksi at Production Wing to 11.47ksi at Production Flowline;

68 Choke and Manifold Tie-in, as per Statoil Design Requirement, ref /8A/.

69 The TH sealing stack features MTM seals with a temperature rating of 177° C and a minimized path of potential leakages.

70 The maximum depth of 500m is given by the actuators of the XMT valves.

71 The chosen materials of the XMT are compliant with ISO 13628/ 10423 material class HH on the Production and class EE for the annulus side.

72 The tree stalk provides landing, locking and sealing for 5 ½” production tubing.

73 It provides a pressure-containing conduit connection the tubing hanger bore to the production flow line through the spool body outlet.

74 The Production Master Valve (PMV), the Annulus Master Valve (AMV) and the Work Over Valve (WOV) are integrated in XT spool body.

75 In the base case configuration of the Morvin Project, i.e. a XMT design with the functionality of the Kristin design, the Production Wing Valve (PWV) and the Cross Over Valve A (XOV-A) are integrated in the spool-mounted Production Wing Block (PWB).

…

Figure 5.9 Screenshot depicting a subset of a semantically annotated excerpt from a document

Table 5.7 Screenshot of a subset of Anchored Document Versioning.

132	**1. General requirements**	**1. General requirements**
133	1.1. General	1.1. General
134	1 Reliability for units and availability for the total Subsea Production Control System for the lifetime of the field shall be documented prior to fabrication.	1 Reliability for units and availability for the total Subsea Production Control System for the lifetime of the field shall be documented prior to fabrication.
135	2 MTTF for subsea retrievable units shall be equal or higher than the lifetime of the field. MTTF in this context is referring to failure causing production loss.	2 MTTF for subsea retrievable units shall be equal or higher than the lifetime of the field. MTTF in this context is referring to failure causing production loss.
136	3 Field proven equipment and techniques should be used.	3 Field proven equipment and techniques should be used.
137	4 Any equipment that requires further qualification and/or verification testing shall be according to WR1622 and procedures approved by Company.	4 Any equipment that requires further qualification and/or verification testing shall be according to CD&E01 Qualify Technology and procedures approved by Company.
138	5 The communication system and subsea power distribution system shall cater for various populations of installed slots on the manifold.	5 The communication system and subsea power distribution system shall cater for various populations of installed slots on the manifold.
139	6 If one umbilical is used, a maximum of 12 wells should be interconnected and controlled from the same umbilical.	6 If one umbilical is used, a maximum of 12 wells should be interconnected and controlled from the same umbilical.
140		7 Contractor shall utilize a FRACAS system including
141		a searchable product test results
142		b and searchable failure data from the complete product lifecycle
143		c or start an improvement programme with the plan to achieve the above in agreement with Company
144	Software product requirements	Software product requirements
145	1 All data generated by the subsea system shall be routed to SCU or other units connected to technical network on host facility.	1 All data generated by the subsea system shall be routed to SCU or other units connected to technical network on host facility.
146	2 Access for data from remote applications on office net or intranet/internet should be via the remote Access@Plant solution described in TR1658.	2 The subsea system shall have facilities for access from shore for monitoring and maintenance of the system. Access for data from remote applications on office net or intranet/internet should be via the remote Access@Plant solution described in TR1658.

6. Conclusion and possible directions

The research applied a semantically oriented approach to automate the retrieval of relevant inferences involved in requirements analysis (RA) for a subsea oil extraction project. In this exploration, an ontological model that expresses the set of concepts involved in the RA phase was generated in an automatic-supervised mode. Having established the ontological model, the benefits that an ontological organization of the domain knowledge provides to text management (or other unstructured information) during the RA phase were quantitatively substantiated. These benefits arise from the direct availability of semantic notions within the texts (e.g. design specification themselves) that support several useful inferences, such as the automatic tracing of requirements insisting on the same components, or the alignment of requirements insisting on similar physical properties as well as the clustering of similar, but possibly inconsistent, requirements spread throughout large documents. The information is presented to the engineer/analyst to assist the design decision process.

In this research, the effectiveness of natural language processing (NLP) and machine learning (ML) techniques in automatizing semantic annotation of requirement documents is demonstrated by locating sentences expressing requirements and assigning them specific (ontological) types. These types refer here to the underlying ontological model but, given the adoption of web-based standards, they are significantly reusable across projects and scenarios. An extensive pilot in the subsea oil extraction domain was performed and quantitative benchmarks are reported. Future directions are also envisioned in view of a viable new generation of RA systems where text retrieval, semantic inference and predictive analysis (about the resulting engineered system) are integrated.

Application of the technologies and processes pioneered by this study to the oil and gas (O&G) industry will provide incrementally more sophisticated processing capabilities. The methodology proposed to design, develop, deploy and validate a semantic requirement extraction system will be able to support the main stages of requirement management services and to enable strictly semantics-driven search and navigation processes. Given the Norwegian subsea contractor (NSC) documentation, once the requirements were recognized, extracted and translated into a machine-readable form, the system could support analysts in the following range of activities:

- Uploading the requirements into third-party requirement management systems, (such as IBM Doors, Vitech Core, etc.);
- Retrieving the requirements through ad-hoc query languages to improve the discovery of phenomena that would not have been made available by standard information retrieval technologies;
- Concept-driven versioning of the documentation; nb: the technology of versioning that has been developed by Reveal srl for technical documents (e.g. Design Basis) will be applied to the conceptualizations derived by the system, improving the robustness of the project-wide alignment of documentation over the lifecycle of the system.

Although the adopted models and the experiments carried out in the investigation described in the above sections cannot be considered definitive, a set of early results can be described as follows:

- The explicit representation of ontological notions, such as all domain concepts, relations and properties, has a strong impact on the accuracy and costs of the project engineering. Since this work falls in the early stages of requirements management, such representation is the foundation for the use of any software platform for supporting requirement analysis. Ontological knowledge is implicit in the analyst work on text documentation. Ontological knowledge is exploited in this work as a building block for the automation of a number of necessary operations, such as locating the proper information (e.g. reference to ISO standard documents) or capturing the relationship between a requirement and the device responsible to satisfy it. The present implementation of ontological modelling is effective in supporting, speeding up and enhancing the analysts' processing and improve their productivity and accuracy.
- The role of textual information is critical as it is the source of ontological information (such as the terminology introduced by the ISO norms). It is also informative of the nature of a target project (e.g. the response to a tender), as it highlights its aims, constraints and requirements.
- Extracting conceptually relevant information from the text involved in RA is a complex task that has been effectively automatized by the extensive use of inductive technologies and support vectors machines (SVM).
- SVMs enabled a direct application of the ontology model and was able to automatize the recognition of the majority of the semantic properties declared in the ontology with incoming free format input text (e.g. a project design document as well as a tender), so that new texts are harmonized with previous information/knowledge whatever its source. Additionally, all concepts recognized or assigned to a text through classification provide a relevant body of evidence (semantic metadata) about texts' content, so that effective indexing and retrieval is enabled for different stages of the RA process.
- Experimental results carried out on a medium sized annotated data sets (i.e. examples used to train our SVM models) show how the semantic annotation task can be automatized with great accuracy at different levels of the RA analysis. Baseline performance measures (with lower baselines understandingly achieved when dealing

with deeper levels of the ontological hierarchy, i.e. at a finer grain and consequent lower statistical frequency) are significantly improved in all the processed texts.

- The relatively short time required to develop a semantic annotation tool for documents entering the RA chain is a clear demonstration of the potential of these methods in realistic operational RA settings.

In summary, the pilot project has been successful in applying NLP and ML to enable automatic semantic annotation of requirement documents by autonomously identifying and locating sentences expressing requirements and assigning them specific (ontological) types. This allocation of types to instances of requirements is the basis for requirement management through commercially available software. Semantic types used here refer to a basic ontological model and are based on strict adoption of web-based standards. The quantitative benchmarking reported confirms the viability of the approach.

The following future development opportunities are foreseen:

- Extend, consolidate and generalize the ontological model developed so far, by the use of an ontology management tool to speed up the future knowledge management activities.
- Scale up the experimental process, by increasing the size and depth of the adopted training material. This would allow evaluating the training curve of the proposed model so that a more precise idea of the impact of these techniques in the operational settings of the analysts can be made available.
- Augment the set of phenomena with point-wise information extraction functionalities, to automatize the recognition of more specific phenomena, such as fine-grained specific entities or relations of the ontology.
- Collect a comprehensive case study set, able to provide the basis onto which an exhaustive and possibly quantitative evaluation can be derived for the whole industry.
- Involve a larger user community in order to carry out a community oriented evaluation campaign, e.g. a cross-division and cross-competence study.

This page intentionally left blank.

7. References

Theory of Design

DNV-GL, 2011. *DNV-RP-A203 Qualification of New Technology*. Oslo: DNV.

DNV-GL, 2014. *Subsea Facilities - Technology Developments, Incidents and Future Trends* (Report No.: 2014-0113, Rev. 03) (Document No.: 18IM1UH-4). Oslo: DNV.

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. *The Academy of Management Review*, 14 (4), pp. 532-550.

Erikstad, S.O., 2014. Introduction to Marine Systems Design Models and Methods. NTNU, unpublished.

Mason, P., 2006. Evolution of subsea well system technology. *Offshore*. [online] 7 January. Available at: <https://www.offshore-mag.com/articles/print/volume-66/issue-7/subsea/evolution-of-subsea-well-system-technology.html>

Pahl, G., Beitz, W. and Feldhusen, J., Grote, K.-H., 2007. *Engineering Design: A Systematic Approach*. 3rd ed. London: Springer-Verlag.

Ree, M., 2012. Disse prosjektene sprekker med 40 milliarder. *Teknisk Ukeblad*, [online] 27 August. Available at: <https://www.tu.no/artikler/disse-prosjektene-sprekker-med-40-milliarder/236641>

Suh, N.P., 1990. *The principles of design*. New York: Oxford University Press.

Whitfield, R. I., Hills, W. and Coates, G., 1999. The Application of Multi-Objective Robust Design Methods in Ship Design. In: *10th International Conference on Computer Applications in Ship Building*, Cambridge, MA; USA, 7-11 June 1999, pp. 251-263.

Subsea oil extraction

International Organization for Standardization, ISO 13628 Petroleum and natural gas industries – Design and operation of subsea production systems , 2005,2015.

- ISO 13628-1:2005 Part 1: General requirements and recommendations
- ISO 13628-2:2006 Part 2: Unbonded flexible pipe systems for subsea and marine applications
- ISO 13628-3:2000 Part 3: Through flowline (TFL) systems
- ISO 13628-4:2010 Part 4: Subsea wellhead and tree equipment
- ISO 13628-5:2009 Part 5: Subsea umbilicals

- ISO 13628-6:2006 Part 6: Subsea production control systems
- ISO 13628-7:2005 Part 7: Completion/workover riser systems
- ISO 13628-8:2002 Part 8: Remotely Operated Vehicle (ROV) interfaces on subsea production systems
- ISO 13628-9:2000 Part 9: Remotely Operated Tool (ROT) intervention systems
- ISO 13628-10:2005 Part 10: Specification for bonded flexible pipe
- ISO 13628-11:2007 Part 11: Flexible pipe systems for subsea and marine applications
- ISO 13628-15:2011 Part 15: Subsea structures and manifolds

International Organization for Standardization ISO 10423:2009 Petroleum and natural gas industries — Drilling and production equipment — Wellhead and christmas tree equipment

Systems Engineering

Bahill, A.T., Gissing, B., 1998. Re-evaluating systems engineering concepts using systems thinking. *IEEE Transaction on Systems, Man and Cybernetics, Part C (Applications and Reviews),* **28** (4), pp. 516-527.

Blanchard, B.S., 1991. *Systems Engineering Management,* New York: Wiley-Interscience.

Blanchard, B.S., Fabrycky, W.J., 1990. *Systems Engineering and Analysis*, 2nd ed. Englewood Cliffs: Prentice Hall.

Institute of Electrical and Electronics Engineers, 1990. *IEEE 610.12-1990: IEEE Standard Glossary of Software Engineering Terminology*. New York: IEEE.

IEEE Computer Society, March 19, 1998, IEEE Std 1362-1998: *IEEE Guide for Information Technology—System Definition—Concept of Operations (ConOps) Document*

International Organization for Standardization, 2005. *ISO 13628-1:2005 Petroleum and natural gas industries -- Design and operation of subsea production systems -- Part 1: General requirements and recommendations)*. Geneva: ISO.

International Organization for Standardization, 2008. *ISO/IEC 15288:2008 Systems and software engineering -- System life cycle processes*. Geneva: ISO.

International Organization for Standardization, 2011. *ISO/IEC TR 24748-2:2011 Systems and software engineering -- Life cycle management -- Part 2: Guide to the application of ISO/IEC 15288 (System life cycle processes)*. Geneva: ISO.

International Organization for Standardization, 2015. *ISO/IEC/IEEE 15288:2015 Systems and software engineering -- System life cycle processes*. Geneva: ISO.

Karim, J., Somers, T.M., Bhattacherjee, A., 2007. The Impact of ERP Implementation on Business Process Outcomes: A Factor-Based Study. *Journal of Management Information Systems,* 24 (1), pp. 101-134.

Seethamraju, R., Diatha, K.S., 2013. Influence of ERP systems on business process agility. *IIMB Management Review*, 25 (3), pp. 137-149.

United States Environmental Protection Agency, https://www.epa.gov/eg/industrial-effluent-guidelines#studies

Valerdi, R., Davidz H.L., 2009. Empirical Research in Systems Engineering: Challenges and Opportunities of a New Frontier. *Systems Engineering - The Journal of the International Council on Systems Engineering (INCOSE)*, 12 (2), pp.169-181.

van den Bosch, A., Bogers T. and de Kunder M., 2016. Estimating search engine index size variability: a 9-year longitudinal study . *Scientometrics,* 107 (2), pp. 839-856.

van El, C.G. et al. on behalf of the ESHG Public and Professional Policy Committee, 2013. Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics. *European Journal of Human Genetics,* 21(Suppl 1), pp. S1–S5.

INCOSE(2015). Systems engineering handbook: A guide for system life cycle processes and activities. (4th ed.) D. D. Walden, G. J. Roedler, K. Forsberg, R. D. Hamelin, & T. M. Shortell (Eds.). San Diego, CA: International Council on Systems Engineering. Published by John Wiley & Sons, Inc.

Yasseri, S., 2014. Application of systems engineering to subsea development. *Underwater Technology*, 32 (2), pp. 93-109.

"youtube.com Traffic Statistics". Alexa Internet. Amazon.com. *July 9, 2017*. http://www.internetlivestats.com/

Introductory material on Ontologies and Description logic

Abecker, A., van Elst, L., 2004. Ontologies for Knowledge Management. In: Staab, S., Studer, R. eds., 2009. *Handbook on Ontologies*. 2nd ed. Berlin Heidelberg: Springer-Verlag, pp. 713-734.

Baader, F., Nutt, W., 2003. Basic Description Logics. In: Baader, F. et al. eds., 2010. *The Description Logic Handbook: Theory, Implementation and Applications*. 2nd ed. Cambridge, UK: Cambridge University Press, pp. 47-100.

Borgida, A., Brachman, R.J., 2003. Conceptual Modeling with Description Logics. In: Baader, F. et al. eds., 2010. *The Description Logic Handbook: Theory, Implementation and Applications.* 2nd ed. Cambridge, UK: Cambridge University Press, pp. 377-401.

Guarino, N., Oberle, D. and Staab S., 2009. What Is an Ontology? In: Staab, S., Studer, R. eds., 2009. *Handbook on Ontologies*. 2nd ed. Berlin Heidelberg: Springer-Verlag, pp. 1-17.

Horrocks, I., Patel-Schneider, P.F. and McGuinness, D.L., Welty, C.A., 2007. OWL: a Description-Logic-Based Ontology Language for the Semantic Web. In: Baader, F. et al. eds., 2010. *The Description Logic Handbook: Theory, Implementation and Applications.* 2nd ed. Cambridge, UK: Cambridge University Press, pp. 458-486.

Hyvönen, E., 2009. Semantic Portals for Cultural Heritage. In: Staab, S., Studer, R. eds., 2009. *Handbook on Ontologies*. 2nd ed. Berlin Heidelberg: Springer-Verlag, pp. 757-778.

Maedche, A., Staab, S., 2000. Discovering Conceptual Relations from Text. In: ECAI, *14th European Conference on Artificial Intelligence*. Berlin, Germany 20-25 August 2000. Amsterdam: IOS Press, pp. 321-325.

Manning, C.D., Raghavan, P. and Schütze, H., 2008. *An Introduction to Information Retrieval.* New York, NY, USA: Cambridge University Press.

Manning, C and. Schütze, H. 1999. Foundations of statistical natural language processing. MIT Press, Cambridge, MA, US..

https://nlp.stanford.edu/IR-book/html/htmledition/support-vector-machines-and-machine-learning-on-documents-1.html Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press. 2008. ISBN: 0521865719.

Middleton, S.E., De Roure, D. and Shadbolt N.R., 2004. Ontology-Based Recommender Systems. In: Staab, S., Studer, R. eds., 2009. *Handbook on Ontologies*. 2nd ed. Berlin Heidelberg: Springer-Verlag, pp. 779-796.

Nardi, D., Brachman, R.J., 2003. An Introduction to Description Logics. In: Baader, F. et al. eds., 2010. *The Description Logic Handbook: Theory, Implementation and Applications*. 2nd ed. Cambridge, UK: Cambridge University Press, pp. 1-44.

Noy, N.F., McGuinness D.L., 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*. [pdf] Stanford: Stanford University. Available at: <https://protege.stanford.edu/publications/ontology_development/ontology101.pdf>

Stevens, R., Lord, P., 2009. Application of Ontologies in Bioinformatics. In: Staab, S., Studer, R. eds., 2009. *Handbook on Ontologies*. 2nd ed. Berlin Heidelberg: Springer-Verlag, pp. 735-756.

Velardi, P., Missikoff, M. and Basili, R., 2001. Identification of relevant terms to support the construction of domain ontologies. In: HLTKM, *Workshop on Human Language Technology and Knowledge Management - Volume 2001*, Stroudsburg, PA: Association for Computational Linguistics, pp. 5:1–5:8, Abney, S., 1996. Statistical Methods and Linguistics. In: Klavans, J., Resnik P. eds. 1996. *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge, MA: MIT Press, pp. 1-26.

Basili, R. et al., 2012. Innovation-Related Enterprise Semantic Search: The INSEARCH Experience. In: *2012 IEEE Sixth International Conference on Semantic Computing (ICSC).* Palermo, Italy 19-21 September 2012, pp. 194-201.

Basili, R., Cammisa, M. and Moschitti, A., 2005. Effective use of WordNet semantics via kernel-based learning. In: *9th Conference on Computational Natural Language Learning (CoNLL 2005)*, Ann Arbor, MI, USA, 29-30 June 2005, pp. 1-8.

Basili, R., Moschitti, A., 2005. *Automatic Text Categorization: from Information Retrieval to Support Vector Learning*. Ariccia: Aracne.

Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. New York: Springer-Verlag.

Brown, B., Chui, M. and Manyika J., 2011. Are you ready for the era of 'big data'? *McKinsey Quarterly*, [online] October 2011 Available at: <https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data>

Buckley, C., Salton, G. and Allan J., 1993. The SMART Information Retrieval Project. In: *Workshop on Human Language Technology*. Princeton, NJ, USA 21-24 March 1993, p. 392.

Cardie, C., Mooney R.J., 1999. Machine Learning, 34, 5–9 Kluwer Academic Publishers, Boston

Burbidge, R., Buxton, B., 2001. *An Introduction to Support Vector Machines for Data Mining*. [pdf] London: UCL Computer Science Department. Available

at: <http://svms.org/tutorials/BurbidgeBuxton2001.pdf>

Croce, D., Moschitti, A. and Basili, R., 2011. Structured Lexical Similarity via Convolution Kernels on Dependency Trees. In: *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing*. Edinburgh, United Kingdom 27 – 31 July 2011, pp. 1034-1046.

Deerwester, S. et al., 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41 (6), pp. 391-407.

Dowty, D.R., Wall, R.E. and Peters, S., 1981. *Introduction to Montague Semantics*. Dordrecht: D. Reidel Publishing Company.

Garzoli, F. et al., 2012. Robust Requirements Analysis in Complex Systems through Machine Learning. In: *JIMSE - Joint workshop on Intelligent Methods for Software System Engineering*. Montpellier, France 28 August 2012, pp. 38-46.

Giannone, C., Basili, R. and Naggar, P., Moschitti, A., 2011. Supervised semantic relation mining from linguistically noisy text documents. *International Journal on Document Analysis and Recognition (IJDAR)*, 14 (2), pp. 213-228.

Gruber Thomas R., 1993.A Translation Approach to Portable Ontology Specifications.*Knowledge Acquisition*, 5(2):199-220,.

Habib, M.S., 2008. Addressing Scalability Issues of Named Entity Recognition Using Multi-Class Support Vector Machines. *World Academy of Science, Engineering and Technology, International Science Index 13, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 2 (1), pp. 17 - 26.

Hadi, W.M., Thabtah F. and Abdel-jaber H., 2007. A Comparative Study using Vector Space Model with K-Nearest Neighbor on Text. In: *World Congress on Engineering 2007 Vol. I*. London, United Kingdom 2-4 July 2007, pp. 296-300.

Jurafsky, D. et al., 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Lenci, A. 2008. Distributional semantics in linguistic and cognitive research. In: Lenci A. ed., From context to meaning: distributional models of the lexicon in linguistics and cognitive science. *Italian Journal of Linguistics*, 20 (1), pp. 1-31.

Liang, A., Salokhe, G., Sini, M., Keizer, J., 2006. Towards an Infrastructure for Semantic Applications: Methodologies for Semantic Integration of Heterogeneous Resources. FAO Cataloging & Classification Quarterly.:43:161–189

Manning, C.D., Raghavan, P. and Schütze, H., 2008. *An Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press.

Mitchell, T. 1997. *Machine Learning*. New York, NY, USA: McGraw-Hill.

Moschitti, A., Pighin, D. and Basili, R., 2008. Tree Kernels for Semantic Role Labeling. *Computational Linguistics*, 34 (2), pp. 193-224.

Moschitti, A., Quarteroni, S. and Basili, R., Manandhar, S., 2007. Exploiting Syntactic and Shallow Semantic Kernels for Question Answer Classification. In: *45th Annual Meeting of the Association for Computational Linguistics*. Prague, Czech Republic 23-30 June 2007, pp. 776-783.

Piciaccia, L.A., 2012. *Ontology Based Project Support*. unpublished, Oslo.

Rifkin, R., Klautau, A., 2004. In Defense of One-Vs-All Classification. *Journal of Machine Learning Research*, 5, pp. 101-141.

SAG, Semantic Analytics Group at the University of Rome, Tor Vergata 2018 http://sag.art.uniroma2.it/topics/machine-learning-2/

Sahlgren, M., 2006. *The Word-Space Model. Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Ph. D. Stockholm University, Department of Linguistics, Computational Linguistics, Stockholm, Sweden

Salton, G., Wong, A. and Yang, C.S., 1975. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18 (11), pp. 613-620.

Sch¨utze, H., & Pedersen, J.(1995). Information retrieval based on word senses.17 In Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval (pp. 161–175).

Sch¨utze, H., & Pedersen, J. (1997). A co-occurrence-based thesaurus and two applications to information retrieval. Information Processing and Management,33(3), 307–318

Shannon, C.E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27 (3), pp. 379–423.

Steinberger, J., Ježek, K., 2004. Using Latent Semantic Analysis in Text Summarization and Summary Evaluation. In: *7th International Conference on Information Systems Implementation and Modelling (ISIM)*. Rožňov p.Radhoštěm, Czech Republic 19-22 April 2004, pp. 93-100.

Vapnik, V.N., 1998. *Statistical Learning Theory*. New York: Wiley-Interscience.

Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag.

Case Study research:

Abercrombie, N., Hill, S. and Turner, B.S., 1984. *Dictionary of Sociology*. Harmondsworth, UK: Penguin.

Campbell, D.T., 1975. Degrees of Freedom and the Case Study. *Comparative Political Studies*, 8 (1), pp. 178-191.

Campbell, D.T., Stanley, J.C., 1966. *Experimental and Quasi-Experimental Designs for Research*. Boston, MA, USA: Houghton Mifflin.

Eckstein, H., 1975. Case Study and Theory in Political Science. In: Greenstein, F.I., Polsby, N.W. eds., 1975. *Handbook of Political Science, Volume 7: Strategies of Inquiry*. Reading, MA, USA: Addison-Wesley, pp. 79-137.

Flyvbjerg, B., 1998a. *Rationality and Power: Democracy in Practice*. Chicago, IL, USA: The University of Chicago Press.

Flyvbjerg, B., 1998b. Habermas and Foucault: Thinkers for Civil Society? *British Journal of Sociology*, 49 (2), pp. 208-233.

Flyvbjerg, B., 2001. *Making Social Science Matter: Why Social Inquiry Fails and How It Can Succeed Again*. Cambridge, UK: Cambridge University Press.

Flyvbjerg, B., 2002. Bringing Power to Planning Research: One Researcher's Praxis Story. *Journal of Planning Education and Research*, 21 (4), pp. 353-366.

Flyvbjerg, B., 2006. Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, 12 (2), pp. 219-245.

Geertz, C., 1995. *After the Fact: Two Countries, Four Decades, One Anthropologist*. Cambridge, MA, USA: Harvard University Press.

Griffin, L.J., Botsko, C. and Wahl, A.-M., Isaac, L.W., 1991. Theoretical Generality, Case Particularity: Qualitative Comparative Analysis of Trade Union Growth and Decline. In: Ragin, C.C. ed., 1991. *Issues and Alternatives in Comparative Social Research*. Leiden, NL: E.J. Brill, pp. 110-136.

Ragin, C.C., 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA, USA: University of California Press.

Ragin, C.C., 1992. "Casing" and the Process of Social Inquiry. In: Ragin, C.C., Becker, H.S. eds., 1992. *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge, UK: Cambridge University Press, pp. 217-226.

Ragin, C.C., Becker, H.S., 1992. *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge, UK: Cambridge University Press.

Uschold, M, Gruninger,M., 2004. Ontologies and semantics for seamless connectivity.ACM SIGMOD Record Homepage archive. Volume 33 Issue 4, December 2004 ACM New York, NY. pp. 58-64

Walton, J., 1992. Making the Theoretical Case. In: Ragin, C.C., Becker, H.S. eds., 1992. *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge, UK: Cambridge University Press, pp. 121-137.

Wieviorka, M., 1992. Case Studies: History or Sociology? In: Ragin, C.C., Becker, H.S. eds., 1992. *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge, UK: Cambridge University Press, pp. 159-172.

8. Published articles and appendixes

SPE-188583-MS

The case for requirement management: How industrial design requirements are specified and executed in the Oil Subsea Production Systems (SPS) vs. Shipbuilding industries

Luca A. Piciaccia,

This paper was prepared for presentation at the Abu Dhabi International Petroleum Exhibition & Conference held in Abu Dhabi, UAE, 13-16 November 2017.

Abstract

None of the recent FPSO projects was delivered on time or on budget. This paper, part of a PhD in Systems Engineering at the Norwegian University of Science and Technology (NTNU), endeavors to make visible some Root Causes by comparing the design process in two industries enabling the upstream oil energy value chain, Subsea oil Production Systems (SPS) and shipbuilding. The comparison uses Systems Engineering concepts and techniques to illustrate the current practices and their reliance on different domain specific conventions.

Particular attention is given to possible commonalities and synergies in the design process to enable a more integrated approach to the industry needs. The purpose is to enable seamless requirement transfer between the user of the vessel and the shipbuilder.

The comparison of the two design processes is leveraging on the 25 years subsea oil production experience of the writer and the guidelines established by professor Stein Ove Erikstad at the Institute of Marine Technology at NTNU in his course “Introduction to Marine Systems Design Models and Methods”.

This paper is not included due to copyright

Subsea Oil Production: The Design Basis

Luca A. Piciaccia, lucapiciaccia@yahoo.com

■ ABSTRACT

The origination of the design basis document for subsea production systems (SPS) in the oil industry is a challenging task. The completeness and correctness of the design basis is crucial for the successful execution of the project in safety, technical, and commercial terms.

This work aims to establish, through systems engineering and through information models (model-based systems engineering (MBSE)), how the SPS design basis should be optimally developed. The MBSE is integrated into this work by the theory of transactional processing, which is implicit in the universal thought process when using behavioral models but seldom explicitly referred to.

The current industry practice for developing an SPS design basis is compared to the suggested optimal process.

The work constitutes the foundation for the revised design basis issuing procedure and the design review practice for the Aker Solutions worldwide subsea business.

This work is part of a PhD at the Norwegian University of Science and Technology.

1. INTRODUCTION

This paper collates direct observations of the design basis generation process across more than two decades and some projects, as implemented in the subsea oil and gas industry by major subsea engineering procurement and construction (EPC) contractors.

The scope of the complete subsea systems engineering value chain is extremely extensive; as a reference, the project CAPEX (capital expenditure) value is typically in the region of USD 50,000.00 to USD 400,000.00 within a field development of up to USD 3,000,000.00.

The focus of this report is therefore limited to one of the most critical processes, the improvements to the method of originating the subsea system design basis in response to the basis of design document issued by the client (operating oil company).

The design basis origination process offers, due to its being a pillar of the overall systems engineering implementation and its potential of developing into a single point of failure in the subsea oilfield development systems engineering process.

It can be added that, in the subsea oil industry as in many others, the design basis is also pivotal to change management and variation impact assessment.

Engineers perform an analysis of the observed deviations from the systems engineering mainstream practice and describe a proposal for a corrective action implementation. The work will constitute the foundation for the revised design basis issuing procedure and the design review practice for the Aker Solutions worldwide subsea business.

For clarity and convenience, the theoretical basis is located in dedicated chapters (systems engineering – chapters 7 and 8; transactional theory – chapter 9). This layout was the choice since the main target audience for this report is made of experienced systems engineering professionals who do not need an introduction to the subject.

The foundations of this work are the systems engineering theory and the transactional theory. The novelty is the explicit reference to the transactional theory; its introduction occurred in response to the peculiarity of the task.

There is a possible tautology in using systems engineering processes for analysing another systems engineering process, the design basis generation. To prevent any self-referencing result, the author found the transactional theory most useful in ensuring consistent and repeatable analysis.

The transactional theory is widely used as a foundation in database management, to ensure consistency and stability of results. It is particularly useful for the design basis management since one organizes the needs, requirements, stakeholders, components, inputs, and outputs in some sort of database.

The author has 20+ years of direct experience in subsea systems engineering management, providing a unique platform for the observation of how the process of establishing the design basis implements in practice.

2. CHAPTERS ORGANIZATION AND DESCRIPTION

Chapter 2 frames the specific problem that will be the subject of the work.

Chapter 3 introduces the approach taken in executing the work and the choice of the methodologies applied.

Chapter 4 illustrates briefly the benefits achieved through this work and the planned developments that will use this work as starting point.

Chapter 5 informs on the choice of the software tool that will support the work.

Chapter 6 constitutes the application of the systems engineering methods to the actual problem.

Chapters 7, 8, and 9 illustrate the theoretical foundations applied throughout the work.

3. NEEDS IDENTIFICATION, PROBLEM DEFINITION, AND PROBLEM STATEMENT

3.1 Problem Characterization

The generation of the design basis document for subsea production systems (SPS) in the oil industry is a challenging task. The completeness and correctness of this document is crucial for the successful execution of the project in safety, technical, and commercial terms.

The design basis is the document that describes the solution that the EPC (engineering procurement and construction) contractor will realize to satisfy the company's (oil company operator of the oil field) needs as expressed by the company's "basis of design."

The correctly drafted design basis is key to a successful design freeze, which in turn is the primary risk driver for project success.

In the subsea industry, the time available from tender opening to contract award is usually between 30 and 90 days, the time from contract award and design freeze should not exceed 90 days after contract award.

This is an extremely short time to handle the approximately 10,000 original, implicit, derived, and conflicting requirements (reference section 7.4).

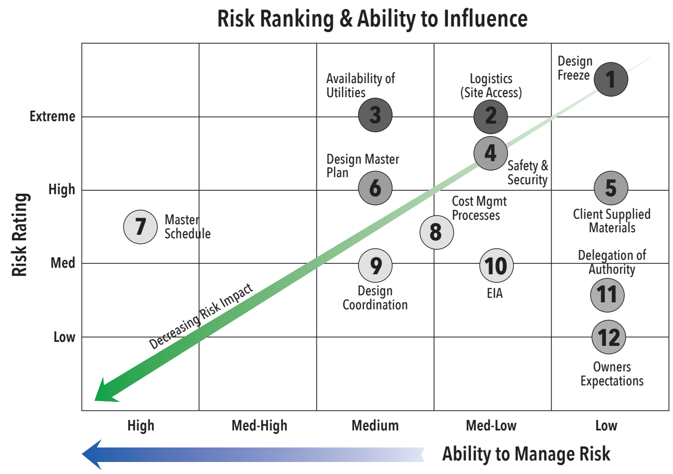


Figure 1. Risk ranking and ability to influence: Derek A. Price, Parsons Brinckerhoff presented at INCOSE 2004

Table 1. Risk ranking

Sum Risk Score	
Intolerable level	Red: Unacceptable Risk Level – Mandatory mitigtion. Risk cannot be justified except in extraordinary circumstances
ALARP high	Orange: ALARP Level – Action required. Risk Tolerable only if further risk reduction is impracticable or if its cost is grossly disproportionate to the improvement gained
ALARP medium	Yellow: ALARP Level – Action required. Tolerable risk if the cost of risk reduction would exceed the improvement gained
ALARP low	Green: ALARP Level – Action may be suggested to mitigate risk. It is necessary to maintain assurance that risk will not excalate from this level
Acceptable level	Green: Acceptable risk

3.2 Problem Relevance

Failure to achieve the "design freeze," the final agreement on what is to be built and delivered to satisfy the stakeholders' needs, is the number one risk factor regarding consequences and the least recoverable of all of the project's possible miscalculations.

The probability of this risk to be actual, that is that the adverse consequence does indeed materialize, can be assumed at 100%.

It is evident that in the classical risk management definition (risk = probability x consequence) any event of missed design freeze is automatically ranked as "critical" (intolerable).

3.3 State of the Art in the Industry

A large body of knowledge exists on the generation of requirements and the recommended way to ensure their verification and validation.

The process of requirements elicitation is an establishment in the discipline of systems engineering, and there are a small number of highly developed commercial software tools that can support the effort system characterization. Notable examples of such tools are CORE™ from Vitech Corporation, "Teamcenter" from Siemens, and DOORS (Dynamic Object-Oriented Requirements System) from IBM. The main question is then: why are these tools not embedded in the subsea oil industry processes?

One common characteristic of mainstream systems engineering is the bias towards programs rather than projects. The main distinction here is the time frame in which such endeavors are carried out. A program typically spans from 5 to 20 years, in the subsea oil industry a project has a delivery cycle seldom longer than two years.

To base this work on the solid foundation, we conducted informal research among the subsea system engineering department senior lead engineers/systems engineering managers to determine which practical hurdles prevent the adoption in full of systems engineering principles.

The results of this informal research (based on the approx. 300 man-years of specific work experience collectively possessed by the interviewees) are summarized.

The consensus is that the existing general practice in systems engineering outside of the subsea industry does not account for:

3.3.1 Technical and Physical Characteristics of a Large Integrated Sub-Sea Oil Development

The subsea oil production field is not fully characterized from the beginning. After the deep subsurface geological investigation has discovered a promising target, exploration wells are drilled in a very limited number to assess if the hydrocarbon deposits are of sufficient quality and quantity to warrant the development. This information is based on complex 3D modeling with a high degree of uncertainty. If the model delivers promising results, the project is sanctioned and the subsea hardware tendered. At this stage, the uncertainty on important design parameters is still high. Essential parameters like fluid composition and temperature, gas volume fraction, and geotechnical properties of the sea floor are very often still only sketchy. But the equipment is now being engineered. This poses challenges that, instead of forming the basis for a research effort for which no time is available, are addressed by the hardware configuration flexibility.

3.3.2 Specific End Product and Constrains of the Subsea Oil and Gas Industry

The usual life span of a subsea system is approximately 20 years, and the legacy support infrastructure may well exceed a useful life of 30 years. In this lifecycle, the applicable regulations and standards evolve, often in a disorganized process.

This is a challenge in designing systems that have to abide by several often conflicting regulations. To date no known organized repository enabling systematic cross-checking of the applicable regulations exists.

Customers have individually developed over time hundreds of proprietary legacy procedures for which, again, only natural language versions are available, and the documentation is not completely revision controlled or reviewed for obsolescence.

3.3.3 Operational Aspects and Experiences

Comparison of the subsea and offshore industry to aerospace yields a notable difference.

An airplane handling system mandates specific interfaces and crossover tools where required that are pre-defined and repeated in all locations where the airplane shall be handled. Airplane handling occurs in a controlled environment under specified conditions. Even transport of components or subsystems has dedicated ad-hoc vehicles.

The subsea equipment has to interface with vessels-of-opportunity (the commercial designation of a vessel chartered for a 1-off campaign) which are varying in performance and handling equipment. More important, these vessels are confirmed, and their detailed characteristic made known, much later than the contract award. This renders the traditional systems engineering approach overly problematic.

3.3.4 Extensive Use of Specific Technical and Economic Legacy Information

The heavy dependence on legacy data for which no structured requirement analysis exists and whose technical qualification of is based on a previous successful use in the early phases of the subsea industry (field was proven) poses a particular challenge to a full requirement verification and traceability.

4. THE PURPOSE OF THIS WORK

This work aims to provide a usable design basis creation functional behavioural model using FFBD: functional flow block diagram (constructing stimulus-response behavior). This will describe (also graphically) how the process of generating a design basis is to be performed. The FFBD theory is described in section 8.2.

The FFBD use is an effective guidance in the lead engineer's effort of establishing a congruent design basis and is intended to constitute the foundation of procedures and work instructions for the worldwide operations of the company.

The availability of the FFBD, even if at this very early stage, is improving the quality of the training of new lead engineers and engineering managers.

To achieve the expected results, two essential steps have been performed:

1. The complete and explicit count and identification of the entities (performers of activities)
2. The correct and complete allocation of functions to entities (actions to action owners) to ensure complete coverage of the necessary work through sufficient activity decomposition and committed resources.

This clarification of the roles and responsibilities is necessary for a successful decomposition of the otherwise unmanageable large number of tasks originating from the design basis work.

Nevertheless, a formal detailed assignment of task responsibility is often neglected, leaving the lead engineers and product engineers to rely on their experience and holistic assessment of the work, resulting in various degrees of quality and completeness.

This work proposes a roadmap that is central in the preparation of the worldwide procedure for issuing the project design basis and to define the functions and departments (components in the FFBD) roles and relationships.

5. EXPERIENCED AND EXPECTED BENEFITS OF APPLICATION

The work is already contributing to reducing role uncertainty in the company and is facilitating the preparation of higher quality design bases.

Due to the difference in cycle time of the work vs. the projects in which it is applied, six months for this report vs. 24 months of project cycle, a full PDCA (plan, do, check, act) approach has not been possible, since it requires at least a full project cycle for the first iteration. Effectiveness measurements will require at least 24 months for the first confirmed data to be available.

Monthly peer reviews have nevertheless confirmed the relevance of the work output towards the perceived improvement needs. The work has undergone interdepartmental reviews and will be the basis for further company programs.

Some of the measures of effectiveness (MOE), which will be followed up, are (details in section 7.3):

- Complete characterization of the stakeholder, process, inputs, and outputs
- Implementation feasibility within six months
- Reduction of internally induced changes on the overall project
- Reduction of Specification shortcoming induced changes
- Reduction of design freeze cycle time
- Customer satisfaction as number of claims submitted.

6. CHOICE OF SUPPORTING INFORMATION TECHNOLOGY ENVIRONMENT

Since the very first phases, it was evident that support from specific systems engineering software would have been essential for a timely and proper execution of the work.

The complexity of the FFBD and the sheer number of entities identified by the first decomposition would not be adequately handled by standard graphic programs like Visio or Powerpoint.

The software of choice was "CORE University Edition" from Vitech Corporation for its proven track record and availability to the University.

7. THE SIX STEPS OF THE SYSTEMS ENGINEERING PROCESS IN THIS WORK

7.1 Step 1: Identify needs; MOE

This step ensured that all the needs had been converted to a format that can be processed by an information system; the process is described here:

7.1.1 Identification and Classification of Information

The information-gathering phase is crucial in properly establishing needs and requirements. The elicitation phase has to be structured to bring to light the real drivers that have set in motion and regulated the process leading to the formulation of the basis of design and the corresponding contractor's design basis.

7.1.1.1 Source Documents and Lower Tiers of Documents

The typical set of documents that constitute the technical part of a subsea EPC contract has been identified by comparing the most recent relevant tenders. This involved interviewing each lead engineer to elicit the key drivers of each contract and customer.

To maintain the general validity of this work references are here made to classes or types of documents instead of specific ones.

7.1.1.2 Requirements

We did an analysis on several of the main specifications and customers originated design basis to identify the individual requirements. Similar work at master thesis level in system engineering (T. Tårneby master's thesis on Customer Requirements for a Subsea Production Control System, Aker Solutions) integrated this as well. It was soon evident that addressing the hundreds of requirements that even a cursory scan identified was not possible in the available period. We then introduced the use of requirement classes. As stated in section 7.4.6 it is possible to work at the requirement class level without adverse consequences for the usability of the conclusions. Once you establish the process, its application to a larger number of individual requirements is an exercise in diligence that does not alter the foundation of the method.

7.1.1.3 Functions

We redid functional mapping from scratch to establish the intended ideal process to be compared with the actual one. This approach is similar to the transactional lean 6 - sigma analysis with the simplifications necessary to fit the schedule. The most important simplification is that no full lean event or Shingi occured, but that we collected the information in sequential interviews.

7.1.1.4 Input and Output

We identified a workable set of input and outputs, again with the help of classes of information rather than specific individual requirements.

7.1.1.5 Stakeholders

We conducted a full stakeholder identification effort via the peer review described in Section 7.6. Once again, we used classes of stakeholders instead of specifics.

7.1.1.6 Components

We identified the entities performing the actual actions again from scratch through the company PEM (project execution manual), the company work instructions and organizational subdivisions as reported by the organization's quality manual.

7.1.1.7 Context Systems

Since the system of interest is the creation of a design basis, it would seem logical to describe clearly the boundaries of the systems and declare as context systems that lie outside of said boundary, which each engineer seems to perceive as a convex Venn diagram. (A convex diagram allows connecting 2 points with a straight-line non-exiting the diagram itself).

Interesting enough, the border of the context seems at times blurred by years of legacy behaviours, which integrate some customers in the fabric of the company by joint research developments. This led to the proposed context system definition to coincide with the component list of Section 7.1.2, which might need further refinement.

7.1.2 System Stakeholders Identification

We identified the system stakeholders and embedded them in the model via component (entities).

Although this work is related to only the "execute" part of the project cycle, we took into consideration both the preceding "win/tender" and subsequent "lifetime services" for the validation of the activities. The definition of the actual components origi-

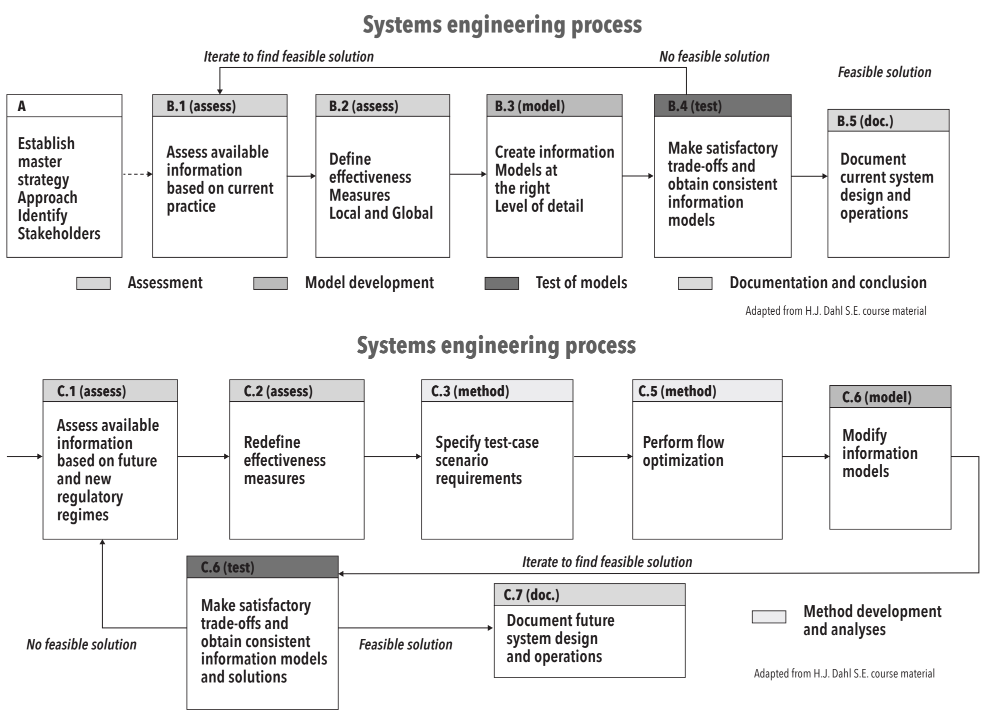


Figure 2. Systems engineering process

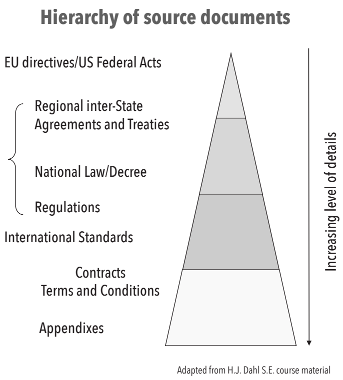


Figure 3 Hierarchy of source documents

Part I – Component List

C *Design Basis Context	int.8 Quality department (QMS)
ext.1 Customer	int.9 R&D departments
int.1 Document Control department	int.10 Sales department
int.2 Field service department	int.11 Sourcing and Logistics department
int.3 Legal Department	int.12 Technical assurance department
int.4 Materials department	proj.1 – Interface coordinator
int.5 Planning Department,	proj.2 – System Engineering Manager
int.6 Product responsible departments	proj.3 – Project Manager
int.7 Project Management	

nated through brainstorming and the lean technique of "fishbone diagrams." We can divide each of the components with a finer granularity, but in their interaction with the project, they all exhibit a "single point of contact" which conveys issues through the component boundary in an orderly manner. This enables us to handle the components as unitized entities.

We uploaded the list of relevant components in the CORE software tool and created the FFBD and the hierarchy diagrams. It is to be noted that we assumed all components abide by transactional behaviour (section 9).

7.2 Step 2: Define Requirements

As pointed out by the systems engineering theory (section 7.4), we classified all requirements along with their origin as original (explicit), implicit, derived, functional, prescriptive, and handled accordingly. Since we deal with classes of requirements and not with individual ones, this process stops, for this work, at the requirements class identification. The further analysis must be performed within each class taking into account the tenets of a proper transactional process as described in Section 9.2 (ACID).

7.2.1 Requirement Traceability

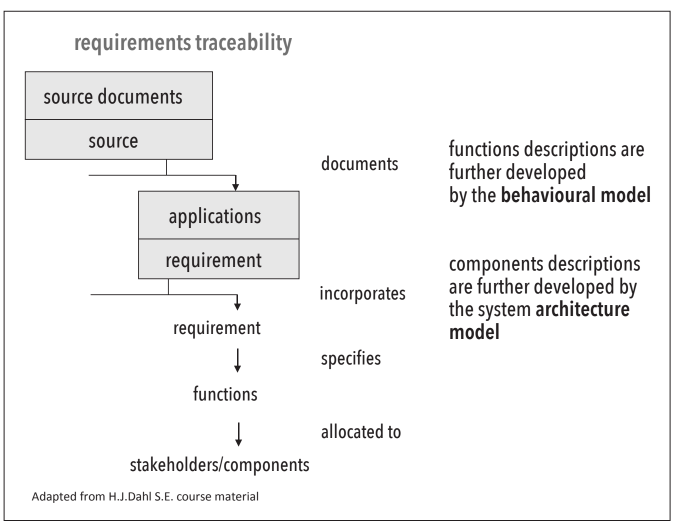
Here, we aim to show the breakdown of requirements from source documents to the final allocation of functions to stakeholders and components.

This is the backbone of the work. Only when the diagram depicted below is true for all requirements and all stakeholders can we state that the project will deliver according to specification.

It is at this stage that the step in section 7.1.1 was brought together to form a cohesive solution to the problem at hand and the behavioural model (section 7.6) has taken shape.

We coded the requirement classes originating from the customer with a prefix CU, the requirement classes originating from the contractor's organization carry no prefix and are intended to convey the full set of requirements from quality management system (QMS), health, safety, and the environment (HSE), manufacturability, supply chain, IP, legal, reliability, availability, maintainability, and more.

The requirement classes are listed in Table 2 on the following page.



7.3 Step 3: Specify Performance

We describe the expected performance and success criteria as headers in the function list. Each posting on the list in real life has a detailed work instruction, procedure or specification document. Where no project specific document occurs, the QMS of the company (ISO 9001 certified) supplies the necessary guidance.

7.3.1 Measures of Effectiveness

The MOEs which we will follow are:

- Complete characterization of the stakeholder, process, inputs, and outputs
  - Achieved by this work and verified by peer reviews
- Implementation feasibility within six months
  - The work resulted in the next review of the design basis procedure with release in March 2012
- Reduction of internally induced changes on the overall project
  - Monitoring of the number, quality, and timing of the change requests will assess differences with previous analogous projects executions, aiming to quantify the benefits of early detection and number reduction of such requests
- Reduction of specification shortcoming induced changes
  - Monitoring of the total number and root causes will quantify the benefits of the new process over project lifecycles
- Reduction of design freeze cycle time

Table 2. Requirements classes

Part I – Requirements Class List

re.1 REQ	from CU Control System Main Requirements / Specification
re.2 REQ	from CU Design Basis for SURF
re.2 REQ	from CU Design Basis requirement document
re.4 REQ	from CU General Basis of Design
re.5 REQ	from CU General Design Requirements for Subsea Systems
re.6 REQ	from CU Meteocean Data
re.7 REQ	from CU Production Profiles and Fluid Characterization
re.8 REQ	from CU Production X-mas Tree System Data Sheet
re.9 REQ	from CU SPS & URF Description and philosophies
re.10 REQ	from CU Subsea Piping for Flow based / Manifolds / SSIV Modules (Pipes, bends, flanges, bolting, gaskets, and more)
re.11 REQ	from CU Subsea Structure general specification
re.12 REQ	from CU WAG X-mas Tree System Data Sheet
re.13 WP01	Management and Administration
re.14 WP02	System Engineering
re.15 WP03	Structures
re.16 WP04	Wellheads
re.17 WP05	XMT
re.18 WP06	Controls
re.19 WP07	Workover
re.20 WP08	Intervention Tooling
re.21 WP10	Tie-in
re.22 WP11	System Testing
re.23 WP20	SLS (Subsea Lifecycle Services)

  - This effect will be the first to be quantified and evaluated for ROI (return on investment) due to the shortest cycle time. Expected feedback 180 days after the introduction of the new design basis generation process.
- Customer satisfaction as a number of claims submitted.
  - This effect will be the last quantified and evaluated for ROI due to the longest cycle time. Expected feedback two years after the introduction.

7.4 Step 4: Analyse and Optimise

7.4.1 Roadmap

This chapter illustrates the approach to the specific challenges of this work and conveys the approach followed in establishing this work roadmap.

7.4.2 Information Models as Basis for the Roadmap

We used information models as the foundation of the work with the objective of providing a method and tool for defining the process, the results and the system.

We completed the following steps, assisted by the tools available in the "CORE university edition" software package:

- Information model notations
  - The notations of CORE underwent adoption even if they differed from the usual notation in the industry; the overall usability of the results was not impaired, and a simple nomenclature re-mapping of few key words is adequate to transpose the results into the specific industrial context.
- Systems Engineering and the "CORE technical" process
  - The "CORE technical" process is the basis for the whole work
  - FFBD: functional flow block diagram (constructing stimulus-response behaviour) is the main vehicle for process description
  - OMT: Object modelling techniques have been consistently used
    - The ERA language
      E: Entity (things: a noun)
      R: Relationship (associations between entities: a verb)
      A: Attributes (define properties: adjective, adverb)
    - Assess available information
      - We carried out a thorough analysis of the information source. A large amount of detailed information required the use of classes of information to undergo processing as individual items. As explained this is not impacting the validity of the results.
    - Define effectiveness measures
      - We proposed effectiveness measures and peer reviews
    - Create information models
      - a: Create requirement traceability model
        - performed via CORE
      - b: Create architecture model
        - performed via CORE
      - c: Create behaviour model
        - performed via CORE
      - d: Create interface model
        - not performed due to time constraints

7.4.3 System Architecture

To perform the necessary actions and foster the appropriate behaviours, the company adopted a sectioned structure coordinated by the PEM and the QMS. To comply with the CORE notations, the departments and actors of the process are components.

The components are interrelated and the arranged in subsystems.

A proper understanding of the structure and operation is graphically described by the hierarchy diagram in Figure 3. The diagram is a standard feature of the CORE software and is derived from the "manual" decomposition of the system as carried out by the system analyst; in this case, the writer extensively used peer reviews for quality and consistency.

7.4.4 Context

The system interacts with surrounding systems; these coincide with the stakeholders indicated in the component list in section 7.1.2. It is important to specify that the interaction is regulated, mediated, and facilitated by the PEM and the QMS. In the

Table 3. Function list

Part I – Function List

1 establish Revision Control (Version History)
2 establish project Framework (Introduction)
2.1 establish Definitions
2.3 establish Project Work Package Build-up
2.4 establish Document Hierarchy
2.4.1 Establish Document order of precedence
2.4.2 Establish Client Specification hierarchy
2.5 establish References
2.5.1 Establish Company References
2.5.2 establish Contractor References
3 establish Field Description
3.1 establish General description
3.2 establish Field layout
3.2.1 establish actual Coordinates
3.2.2 Establish actual Water depth
3.5 establish Fluid Composition
3.6 establish Environmental conditions
3.7 establish Soil conditions
4 establish Design Parameters & prerequisites
4.1 establish Design Water Depth
4.2 establish Design Life
4.3 establish System Availability target
4.4 establish Design Weight
4.5 establish Dropped objects resistance targets
4.6 establish installation philosophy
4.7 establish Overtrawlability
4.8 establish Design Pressures
4.9 establish Design Temperatures
4.10 establish Material Selection
4.11 establish Insulation requirements
4.12 establish ROV intervention design criteria
4.13 establish Tagging requirements
4.14 establish Piping and Wiring requirements
4.14.2 Establish Service Line
4.14.3 Establish Gas Lift
4.14.4 establish Water Alternating Gas (WAG)

4.14.5 establish Water Injection (WI) lines
4.14.6 Establish Gas Injection (GI) lines
4.14.7 Establish Methanol
4.14.8 Establish Hydraulic lines
4.14.9 Establish CHemical lines
4.14.10 establish Electrical lines
5 establish Subsea Production System
5.1 establish Scope of supply
5.2 establish Main deliverable descriptions
5.2.2 Establish Tie-in system
5.2.3 Establish Wellheads
5.2.4 Establish Umbilical
5.2.5 Establish XT
5.2.6 Establish Controls
5.2.6.1 Establish Hydraulic System
5.2.6.2 Establish Hydraulic supply system
5.2.6.3 establish Power and Communication System
5.2.6.4 establish Subsea Control Module
6 establish Design Codes
6.1 establish National regulations
6.2 establish International standards
7 establish Reliability requirements
8 establish Safety requirements
9 establish Type approval requirements/ certification
10 establish Packaging & transport requirements (major items)
11 establish Testing requirements
12 establish Environmental requirements
F.0 Perform Design Basis function
F.1 perform Systems Engineering functions
F.2 summarize customer's Basis of Design
F.3 summarize technical equipment specifications
F.4 define the SCOPE OF WORK
F.5 define DESIGN CODES
F.6 define FIELD DESCRIPTION
F.7 define DESIGN PARAMETERS & PREREQUISITES

case of different legal entities cooperating, such activity is regulated by existing or ad-hoc intercompany agreements. These agreements cover compensation, IP rights, risk apportioning, and management.

7.4.5 System Boundary

The system of interest is the design basis development (process) inside the company; it interfaces with all the stakeholders indicated in the component list Section 7.1.2.

7.4.6 Behavioural Model

The consistent use of behavioural modelling has facilitated the complete and correct list of the functions discharged by the design basis, overall, we identified 72 main functions and allocated them to 17 high-level components.

We identified an initial set of 23 requirement classes (requirement sources). The limited amount of time prevented the full unfolding of these requirement classes into the hundreds of individual line requirements that do constitute the requirements body. This incomplete decomposition is not harming the main objectives of the work as listed below.

1. Establish the complete and explicit count and identification of the entities (performers of activities).
2. Establish the correct and complete allocation of functions to entities (actions to actions owners) to ensure complete coverage of the necessary work through sufficient activity decomposition and committed resources.
3. Clarify the roles, responsibilities, inputs, and outputs (deliverables) of all the entities.

7.5 Step 5: Design and Solve

Since the objective of the work is to establish the process for writing a design basis, it is possible to work at the requirement class level without adverse consequences for the usability of the conclusions. Once the process is established, its application to a larger number of individual requirements is an exercise in diligence that does not alter the foundation of the method.

The previous statement is implicitly based on the properties of transactions as expressed by the transactional theory, namely *atomicity, consistency, isolation, and durability* (definitions at the end of the paper), which we consider true in deterministic environments.

7.6 Step 6: Verify and Test

This work centered on establishing an optimal process for the generation of the design basis. This implies the generation and implementation of new work routines and changes to established work processes.

Anyone with experience in organization restructuring knows this is probably the task that takes longest. Within the time span of this work, no complete implementation was possible.

We carried out the verification by monthly peer reviews confirming the relevance of the work output towards the perceived improvement needs.

Also we regularly reported out to the vice president of "Engineering Execution" and the vice president of "Concepts and Studies" to ensure backing to the activity at the appropriate level.

The work underwent interdepartmental reviews and will be the basis for further company programs.

The study demonstrated the impossibility to execute the classic systems engineering approach in handling the large number of explicit, implicit, and derived requirements in the 90 days usually allotted.

Although we demonstrated the systems engineering principles to be applicable in full, the associated tasks are not fitting the resource and time constraints even when using massive parallelization in activity planning.

Further work will be carried out in the writer's PhD program to identify and apply techniques supporting the natural language to formal requirement conversion, and multiple sources cross-referencing, to deliver an effective tool for a drastic reduction of analysis cycle time and information overflow.

8. THEORETICAL FOUNDATIONS

To properly anchor this paper in established systems engineering practice, the systems engineering theory constituting the foundation of the work are the NTNU Systems Engineering Course EP8102 and the model-based systems engineering as expressed by Oliver, Kelliher, and Keegan in the book *Engineering Complex Systems*.

The MBSE further integrates into this work by the theory of transactional processing, which implicitly relies upon when engineers draft behavioural models, but this seldom comes to light as it is usually taken for granted in the engineer's thought process.

A simplified introduction to the theory follows in the next sections.

8.1 The Systems Engineering Process Model

Any process can be a behaviour. This process involves the steps (transactions) taken, the inputs and outputs for each stage, and the sequencing of the transactions.

One can depict the process with a FFBD. Here, we consider six fundamental modelling steps. We group these under the "CORE" branch of the *systems engineering technical process*, as described in Figure 4 and termed core steps.

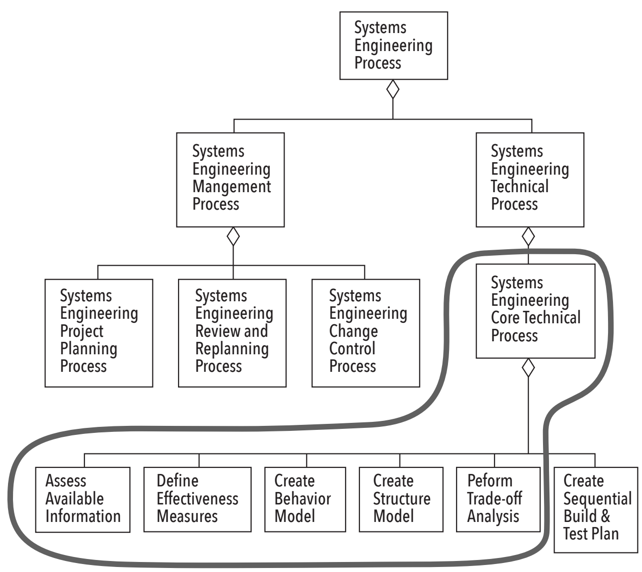


Figure 4 *Extended part list for system engineering process*

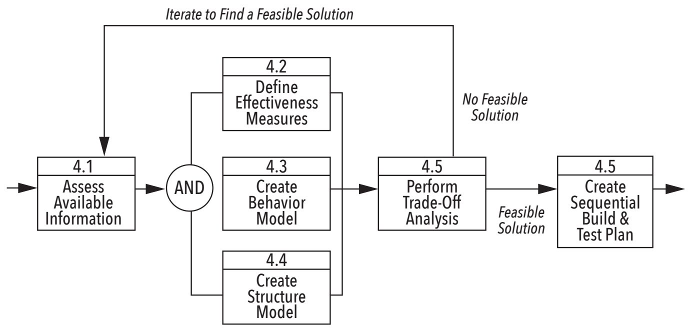


Figure 5. FFBD view for the system engineering core technical process (D. W. Oliver)

The steps considered in this work are those encircled by the red line.

8.2 The "CORE Technical" Process

8.2.1 View of the System Engineering Core Technical Process

The CORE technical process, depicted in Figure 5 via an FFBD to exemplify stimulus-response behaviour, clarifies the order of the six engineering modelling steps that make up the MBSE approach to the systems engineering process.

8.3 The Six Steps in the "CORE Technical" Process

The six steps accomplish the following tasks:

8.3.1 1. Evaluate and Categorize Available Information

Includes the search for missing information.

When the available information has been assessed the next steps in the CORE Process are undertaken.

8.3.2 2. Define the Criteria for Optimization (Measures Of Effectiveness)

These criteria mean success or failure. Effectiveness measures are the small subset of the requirements that are so important that the system will fail if they are not met and will be a success if they are.

They are the criteria used in the trade-off decisions on what to build.

The design of a system is a poorly defined problem that has no feasible solution without a set of criteria to guide choices.

The effectiveness measures might be seen as the equivalent of the regularization functions used in optimal control or calculus of variations.

They are critical because all the stakeholders – engineers, management, users, and operators – must agree on them and the agreement properly documented. Failure to do so results in unacceptable risks.

If these criteria are not both correct and agreed upon, then the system development will be plagued with costly requirements changes and may entirely fail to be "fit for purpose."

The effectiveness measure results are extremely useful in reviews with management, users, and operators who do not want to know all of the technical engineering detail but do want to know how well the system being built will satisfy their needs.

Engineers rank effectiveness measures by a set of priorities established by statistically valid methods (Saaty 1983).

The engineers compute effectiveness measure values from the properties of the system components and their behaviour or established them by group surveys which determine user preferences.

8.3.3 3. Define the Behaviour Desired with an Executable Model

The behaviour model captures what any component or entity has to perform. It shall contain enough information to be executable.

The model must capture all of the steps or functions involved in the behaviour, how the functions interconnect, their expected mode of interaction, and all of the inputs and outputs of each function with the aim to correct all instances of unallocated requirement.

If the interconnection of the components allows alternative responses (paths), then the conditions for the alternative paths must be captured.

8.3.4 4. Define Executable Structure Models

The allocation of behaviour onto Entities (physical or logical) is necessary. All the relevant attributes conductive to performance must be present in the object description.

This CORE engineering step captures the static structure (including all interconnections) of the system context, the subject system, or any meaningful configuration of system components.

Static structure involves the description of components (elements), functions, and their associations.

This information is usually recorded in text and graphically by the use of several established techniques such as the IDEF0, the FFBD, and the N2 notations and diagrams.

8.3.5 5. Perform a Trade-Off Analysis

This phase is conductive to the selection of the optimal alternative design or architecture.

Any design to be feasible must meet all of the performance requirements at system level, that is, pass the validation criteria whilst the best feasible design is selected based on the effectiveness measure values such that it passes the verification test.

This is the optimization process. It is a key best practice in the engineering of complex systems.

The attributes and properties of the components derive the system level performance and effectiveness for the alternatives found in the previous steps.

It is often necessary to iterate the trade-off multiple times to fully address a large number of conflicting demands imposed on the system.

It is in the trade-off analysis that the performance requirements and the effectiveness measures undergo the evaluation at the system level. A useful tool for an effective trade-off is the balanced scorecard, which helps in the quantitative expression and ranking of preferences otherwise difficult to compare.

The objective is not the optimization of individual components, but the optimization of the system.

8.3.6 6. Create a Plan

This step controls the implementation of the activities described and must take into account the business realities.

There are several reasons for considering implementation crucial. In some cases, the resources available and the cycle time dictate the sectioning of the system into stages, completed sequentially.

In some cases, the technical work uncovers opportunities or a need for a partnership that must be examined in parallel with the technical development and incorporated in the overall planning.

In many cases, parts of the development are high risk, and this requires mitigation before the early activities planned to accommodate the risks.

8.4 Requirements Traceability

The most common form for systems engineers to receive their information is in large complex text documents written in natural language integrated by graphs, diagrams, or tables.

Frequently the requirements will be mixed with other non-directly relevant information which must be graded and sorted.

Even after extensive processing, the requirements obtained may be redundant, contradictory, incorrect, incomplete, unverifiable, and poorly written.

The other primary source of information is heritage systems. These often constitute the backbone of the commercial offer or are mandated by the customer as building blocks or interfacing systems. They are typically not documented with a sufficiently rigorous methodology and have little readily usable information available without reverse engineering the existing product.

Initial information received at the beginning of a project is often largely in the form of text, heritage information, user information, text operations concepts, and models.

8.4.1 Original Requirements

The original requirements are those explicitly present in the original documentation and are closely related to another category:

8.4.2 Derived Requirements

As the original requirements are studied and as modelling proceeds, additional requirements, such as derived requirements are identified.

8.4.3 Implied Requirements

These have no precursor in any documentation. They represent omissions in the initial information. When identified and formulated explicitly, they become part of the developed information.

In other industries, they might occur less frequently than the other requirements types, whilst in the subsea oil industry, implied requirements do constitute an important percentage of the total requirement count.

These requirements are often "hidden" in contractual references to standards or client internal procedures not attached to the body of the contract.

Often a reference is made to the "last issue of the standard." This is particularly insidious since, for a project spanning several years, the standard is likely to evolve over the project life span, leaving the contractor exposed to unrecoverable change-derived costs.

8.4.4 Inconsistent Requirements

Inconsistent requirements are quite commonly found in system specifications for large projects. Teams each responsible for different functionality build these specifications and these functionalities usually have conflicting demands that need to be ranked and validated during system design.

It is important to detect conflicting requirements in the earliest possible stage to preserve the system design integrity and avoid making incompatible design decisions for parts of the system.

8.5 Requirement Classification by Use

Text requirements are classified by use so that they can be traced or allocated properly to the correct entities, such as functions or components.

This classification helps the project to monitor completeness and correctness of the model by enabling cross-checking: have all performance allocations been made to components? Do all components perform at least a function?

The following is the typical grouping of requirements by use:

8.5.1 Functional Requirements

State what the system must do and trace to the functions which will accomplish them.

In the models, these functions are performed by in elements and appear in as executable behaviours. They do not prescribe how the system will be built, only on what it shall perform and the relevant acceptance criteria.

8.5.2 Temporal Performance Requirements

Quantify the amount of time system has to respond to the stimulus. These time values are allocated to the relevant functions.

8.5.3 Non-Temporal Performance Requirements

Quantify expectations for properties of the system like cost, weight, size, power consumption, availability, security, and more.

These quantities are allocated to the components, which make up the system. The components are physical or organizational entities and must have attributes that match these quantities.

8.5.4 Interface Requirements

Specify input/output types, limits of flow, and timing at the interfaces between components. The behaviour of the components at the interface must be designed to meet the interface requirement. These requirements are increasingly important due to the supply chain globalization. Industry standards in interfaces are critical. In large complex systems, thousands to tens of thousands of interfaces exist and must be consistent for the system to work.

8.5.5 Prescriptive Requirements

The specification documents often do contain design requirements which predetermine a design choice. It is essential to ascertain with the applicable stakeholder if this requirement is meant to apply to the letter or whether it is a misstatement of a requirement in the form of design. The issue traces to a design requirement and a formal resolution. The design requirement becomes either an imposed constraint which will be followed, one of the other kinds of requirements, or it is formally deleted.

9. A BEHAVIOUR FOR ASSESSING THE AVAILABLE INFORMATION

As we have seen, when a team developing a large complex system receives thousands of pages of available information in the form of natural text, there is a large amount of work to be done to assess the information.

- Properly elicit the requirements.
- Identify and correct the problems in the requirements.
- Classify the kinds of requirements so that they can be properly allocated and traced.
- Establish a system enabling the tracking of the engineering progress.
- Create needed traceability links.

We can describe the process for assessing available information as behaviour with a model.

9.1 Effectiveness Measures

It is necessary to classify effectiveness measures based on the kind of work that must be done to evaluate them.

Calculated with equations from the attributes of the components or subsystems – attributes like weight, cost, power, or reliability.

Calculated from modelling and analysis – modelling of behaviour, simulation of the probability of detection, and more.

Obtained from a survey of the preferences of owners, operators, and users using their choices among solution alternatives.

9.1.1 Priorities among Effectiveness Measures

We can handle complex systems with the same approach as simple ones if we can verify the assumption that there are only a few macro-level effectiveness measures driving them. Further decomposition is necessary during the verification and validation since complex systems exhibit several tiers of decomposition and many linear and non-linear interactions.

Simply listing the individual effectiveness criteria is not sufficient to completely establish the solution to opt for.

9.1.2 Continuous Function Approach

After a quantification effort, we might represent the values by pseudo-continuous functions or their approximation, rather than a discrete set of options. In this case, the solution can be identified by minima in the functions, choosing the local or global minimum that satisfies the complete set of needs.

9.1.3 Single Cost Function Approach

If the effectiveness measures are amenable to being integrated into a single cost function by assigning priorities or weights for each measure, the solution with the lowest value of cost function is then the system of choice. This approach is conceptually similar to a balanced scorecard.

The uniform cost function provides a single number on which to base the selection of the design for use. Scenario analysis enables examining the set of individual effectiveness measures and how they vary with alternatives in a verification of the robustness of the modelling and the solution found.

This selection is a selection based on quality/risk and should be quantified with a proper requirement (even if it has to be a derived one).

9.1.4 Priority Survey Approach

The selection of the effectiveness and their weighting factors both require setting priorities by assessing the opinions of all the major stakeholders: owners, operators, users, management, marketing, customers, and more.

Perform the priority survey that establishes priorities for cost functions.

It shows that the effectiveness measures have priorities, generated by the priority survey.

Also, these effectiveness measures can be collated, ranked, and organized using balanced scorecards.

10. THE TRANSACTIONAL PROCESS

10.1 Definition

Transactional process describes quite accurately the natural thought process the systems engineer is applying to reach the issuing of the design basis.

Based on transactional process where the global process (global transaction) is defined as "perform design functions that satisfy all applicable requirements."

10.1.1 Transaction Processing

Information processing divided into individual, indivisible operations, called transactions is transaction processing. Each transaction must succeed or fail as a complete unit; it cannot remain in an intermediate state. Transaction mandatorily requires acknowledgment to get received as a necessary feedback for accomplishment.

Transaction processing maintains the system in a known, consistent state, by ensuring that any operations carried out on the system that are interdependent are either all completed successfully or all cancelled successfully.

Transaction processing allows multiple individual operations to be linked together automatically as a single, indivisible transaction.

The transaction-processing system ensures that either all operations in a transaction complete without error or none of them do. In this case, the absence of faults in the specification is the metric chosen for establishing the transaction's success

If all operations of a transaction complete successfully, the transaction is committed by the system, and all changes to the database are made permanent; the transaction cannot roll back once this happens. In the context of the design basis, the transaction is committed once the procedure has passed the three stages of review and is officially issued by the document control system.

Transaction processing guards against errors that might leave a transaction partially completed, with the system left in an unknown, inconsistent state. The document review cycle ensures this function.

If the system halts in the middle of a transaction, the transaction processing system guarantees that all operations in any uncommitted (not completely processed) transactions undergo cancellation. In the context of the design basis, this means that the document control is a revision-controlled environment that does not allow indeterminate states in the approval cycle.

Transactions process in a strict chronological order. If transaction n+1 intends to touch the same portion of the data as transaction n, transaction n+1 does not begin until transaction n is committed. Before any transaction is committed, all other transactions affecting the same part of the system must also be committed; there can be no "holes" in the sequence of preceding transactions.

The process of a drafting the subsea design basis is inherently a parallel one. Within each of the parallel branches, the previous statement is correct. Nevertheless, a design basis can issue with "holds" or TBAs in the text, and the relevant parameter definition shortcoming managed through the project risk register. This constitutes a "compensating transaction."

10.1.2 Compensating Transaction

In systems where commit and rollback mechanisms are not available or undesirable, a compensating transaction is often used to undo failed transactions and restore the system to a previous state or evolve it to a stable future state.

10.2 ACID criteria (atomicity, consistency, isolation, durability)

Transaction processing has, among others, these benefits for use in writing a design basis:

- It allows sharing of resources among many users
- It allows work decomposition through a WBS (work breakdown structure)

A transaction is a unit of work that has the following properties:

- **Atomicity:** A transaction should be done or undone completely and unambiguously. In the event of a failure of any operation, effects of all operations that make up the transaction should reverse, and data should roll back to its previous state.
- **Consistency:** A transaction should preserve all the invariant properties (such as integrity constraints) defined on the data. On completion of a successful transaction, the data should be in a consistent state. In other words, a transaction should transform the system from one consistent state to another consistent state. For example, in the case of relational databases, a consistent transaction should preserve all the integrity constraints defined on the data.
- **Isolation:** Each transaction should execute independently of other transactions that may be executing concurrently in the same environment. The effect of executing a set of transactions serially should be the same as that of running them concurrently. This requires two things:
  - During a transaction, an intermediate (possibly inconsistent) state of the data should not undergo exposure to all other transactions.
  - Two concurrent transactions should not be able to operate on the same data. Database management systems usually implement this feature using data locking.
- **Durability:** The effects of a completed transaction should always be persistent.

> continued on page 70

- Hey, T. and A. E. Trefethen. 2003. "The Data Deluge: An e-Science perspective." In *Grid Computing – Making the Global Infrastructure a Reality*, edited by F. Berman, G. Fox, and A. J. G. Hey, Chapter 36, Chichester, UK: John Wiley & Sons, Ltd.
- Kalawsky, R. S., J. O'Brien, et al. 2005. «Improving scientists interaction with complex computational-visualisation environments based on a distributed grid infrastructure.» *Philosophical Transactions of the Royal Society A : Mathematical, Physical & Enginering Sciences* 363 (1833): 1867-1884.
- Korn, S., G. R. Burns, and D. Harrison. 1999. "The Application of Multi-paradigm Simulation Techniques to Manufacturing Processes." *International Journal of Advanced Manufacturing Technology* 15 (12): 869-875.
- Leech, J., J. F. Prins, and J. Hermans. 1996. "SMD: Visual steering of molecular dynamics for protein design." *IEEE Computational Science & Engineering* 3 (4): 38–45.
- Liere, R. v., J. D. Mulder, and J. J. v. Wijk. 1997. "Computational steering." *Future Generation Computer Systems* 12 (5): 441-450.
- Miller, R. B. 1968. "Response time in man-computer conversational transactions." Proceedings AFIPS Fall Joint Computer Conference, San Francisco. US-CA, 9-11 December.
- Mulder, J. D., J. J. van Wijk, and R. v. Liere. 1999. "A survey of computational steering environments." *Future Generation Computer Systems* 15 (1): 119 – 129.
- Newman W. M., and A. S. Taylor. 1999. "Towards a Methodology employing Critical Parameters to deliver Performance Improvements in Interactive Systems." Paper presented at Interact '99 Conference, Edinburgh, UK, 30 August–3 September.
- Parulkar, G. M., J. Bowie, H. Braun, R. Guerin, and D. Stevenson. 1991. "Remote visualization: challenges and opportunities." Proceedings of the 2nd Conference on Visualization '91, San Diego, US-CA, 22-25 October: IEEE Computer Society Press.
- Pickles, S. M., R. J. Blake, B. M. Boghosian, J. M. Brooke, J. Chin, P. E. L. Clarke, P. V. Coveney, N. Gonzalez-Segredo, R. Haines, J. Harting, and M. Jones. 2004. "The TeraGyroid Experiment." Proceedings of the Workshop on Case Studies on Grid Applications at GGF10, Berlin, DE, 13 March.
- Watson, B., N. Walker, W. R. Ribarsky, and V. Spaulding. 1998. "The effects of variation in system responsiveness on user performance in virtual environments." *Human Factors* 40 (3): 403-414.
- Yucesan, E., Y.-C. Luo, C.-H. Chen, and I. Lee. 2001. "Distributed Web-based Simulation Experiments for Optimization." *Simulation Practice and Theory* 9 (1): 73-90.
- Zulch, G., U. Jonsson, and J. Fischer. 2002. "Hierarchical Simulation of Complex Production Systems by Coupling Models." *International Journal of Production Economics* 77 (1): 39-51.

ABOUT THE AUTHOR

Professor Roy S. Kalawsky PhD, MSc., BSc., CEng, FIET, FRSA is director of the Advanced VR Research Centre at Loughborough University. Before joining Loughborough University in 1995 he worked for BAE Systems and was responsible for advanced cockpit research across the Military Aircraft Division. In 1993, he received the Royal Aeronautical Society medal for developing the United Kingdom's first virtual cockpit. Kalawsky joined BAE Systems in 1978 as a systems engineer and researched advanced cockpits for fast jet aircraft.

Piciaccia continued from page 61

These properties, collectively known as ACID properties, guarantee that a transaction is never incomplete, the data is never inconsistent, concurrent transactions are independent, and the effects of a transaction are persistent. ■

REFERENCES

- Bernstein, P. A., and E. Newcomer. 1997. *Principles of Transaction Processing*. Morgan Kaufmann, San Francisco, US-CA.
- Elmagarmid, A. K. (Editor). 1992. *Transaction Models for Advanced Database Applications*. Morgan-Kaufmann, San Francisco, US-CA.
- Gray, J., and A. Reuter. 1993. *Transaction Processing – Concepts and Techniques*. Morgan Kaufmann, San Francisco, US-CA.
- Hans, J. D. Materiel from systems engineering course EP8102, Norwegian University of Science and Technology (NTNU), Trondheim, NO.
- Oliver, D. W., T. P. Kelliher, and J. G. Keegan. 1997. *Engineering Complex Systems*.
- Transactional Theory, 1999. https://www.subbu.org.
- Weikum, G., and G. Vossen. 2002. *Transactional Information Systems: Theory, Algorithms, and the Practice of Concurrency Control and Recovery*. Morgan Kaufmann, San Francisco, US-CA.

ABOUT THE AUTHOR

Luca Abele Piciaccia developed his 25+ years systems engineering career in the subsea oil industry at major oil companies and leading EPC contractors, where he held several positions as engineering manager, commissioning responsible, and department manager before becoming chief engineer, Subsea Systems Engineering for major subsea projects. His education includes nuclear engineering at the Polytechnic University of Milan, BSc in engineering management at CCU, and an MBA in technology management from Chiefly Business School in Australia. Luca has served two terms as president of the Norwegian INCOSE Chapter.

Ontology-driven Semantic Search for Requirement Engineering

L. A. Piciaccia (*), R. Basili, D. Croce, V. Storch (+)
(*) Akersolution, (+) University of Roma, Tor Vergata

Luca A. Piciaccia
Aker Solutions
luca.piciaccia@akersolutions.com

Roberto Basili
University of Roma, Tor Vergata
basili@info.uniroma2.it

Danilo Croce Ph.D.
University of Roma, Tor Vergata
croce@info.uniroma2.it

Valerio Storch
University of Roma, Tor Vergata

Abstract. Requirement elicitation is a challenge in the Subsea Oil extraction industry due to the short project award and execution times that do not allow the complete compilation of a requirement database from the contract documents and appendixes. In this paper, a semantic approach to the automation of relevant inferences involved in the Requirement Analysis (RA) stages is proposed. An ontological model is here proposed to supporting the enrichment of the involved texts along a number of semantic assumptions, aiming to provide a rich explanatory description of the targeted phenomena These general types are used to enable the semantic annotation of texts and sentences within the documentation, i.e. conceptual information useful to characterize design choices (as in specification documents), requirements (design documents) or product descriptions (catalogues). A Machine learning approach is then discussed as a robust and effective solution to annotated texts according to the ontological model.

Introduction

Requirement elicitation is a challenge in the Subsea Oil extraction industry due to the short project award and execution times that do not allow the complete compilation of a requirement database from the contract documents and appendixes. In particular, sustainability needs and requirements, in both the domains of industrial lifecycle and industrial ecology appear not to receive the deserved attention.

Every requirement analysis stage relies on strong assumptions about the nature of the underlying information: terminology and normative statements in requirement documents (e.g. design specifications) as well as components, devices and their parts involved in a system. However, these assumptions are fully implicit in the documents. They are for example labels, acronyms or names for the involved components or titles (i.e. column names) for tables or field names for records. Recent work in semantic technologies demonstrated that machine-readable models of the domain semantics increase the ability of the system to reason about their data. Moreover it allows automatizing most of the inferences involved in an application task.

In this paper, a Semantically Oriented approach aiming to automatize relevant inferences involved in the Requirement Analysis (RA) stages is proposed. In particular, we show how Natural Language Processing and Machine Learning techniques can be used to automatize semantic annotation of requirement documents by locating sentences expressing requirements and assigning them specific (ontological) types. Semantic Types refer here to the underlying Ontological model but, given the adoption of WWW standards, they are portable across domains and systems. An extensive investigation in the Subsea Oil extraction industry domain is discussed and quantitative benchmarking is reported. Future directions are also discussed in view of a viable new generation of RA systems where text retrieval, semantic inference and predictive analysis (about the resulting engineered system) are integrated.

The advantages that an ontological organization of the domain knowledge provides to text management (or other unstructured information) during the RA phase are analyzed. These enhancements arise from the direct availability of semantic notions within the texts (e.g. design specification themselves) supporting several useful inferences: concept driven search; automatic tracing of requirements insisting on the same components; alignment of requirements acting on similar physical properties across full the span of the system.

Natural Language Processing and Domain Ontologies

In the following, the discussion of Semantic Annotation of a Document from the Subsea Oil Project Project is provided. The idea is to enrich texts according to a number of semantic assumptions that will give rise to an overall model that we hereafter call **Ontology**.

The automatic recognition of such semantic phenomena would enable more expressive Search Operations, and in general Management, within usually huge amount of Documentations. As an example the ontological enrichment of texts according to the proposed schema would enable Semantic Queries, such as "*Give me all pressure constraints for Xmas trees*" or "*Give me all components of templates located at 800ft*".

Ontological Modeling of Requirements in the Subsea Domain

In order to reproduce the analysts and expert behavior, an ontological organization of the domain knowledge is required, in order to express all products and processes taken into account in the RA stage.

As discussed by Thomas Gruber in (Gruber, 1995), an ontology is an explicit specification of a conceptualization. From a theoretical and philosophical perspective, an ontology is a systematic account of Existence. For Artificial Intelligence (AI) systems, what "exists" is that which can be represented. When the knowledge of a domain, such as the subsea domain, is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities, e.g. products (as an example the subsea valve block arrangement called in jargon Christmas Tree or XMT) or processes (testing process), in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms.

We assume here that an underlying model is available and it is expressed through more or less complex domain ontology. Given this assumption, we proof the benefits that an ontological organization of the domain knowledge provides to text management (or other unstructured information) during the RA phase. These benefits arise from the availability of semantic notions directly within the texts (e.g. design specification themselves) that support several useful inference: automatic tracing of requirements insisting on the same components; alignment of requirements acting on similar physical properties; clustering of similar, but possibly inconsistent, requirements spread in large documents.

The subsea **Ontology** will include a number of **general types** able to describe **categories of entities** in the reality (e.g. Physical Phenomena vs. Physical Entities, such as the pressure vs. a tube, or Abstractions vs. Activities, such as numbers, percentage vs. filling or moving actions). These (limited set of) general types will enable the enrichment of texts and sentences with their semantic information useful to characterize design choices (as in base document), requirements (basis of document) or product descriptions (Aker catalogues).

Given an ontology expressing all the existing concepts in our target Subsea domain, we discuss in the next section a viable data-driven approach to annotate documents with such knowledge. Let us consider a set of high level concepts for the description of all the **`Type_of_Activity`** that are carried out in Aker Solutions, such as `Control_system_deployment`, `Wellhead-production` and `X-mas_trees_deployment`. Moreover, we can add information about physical entity that provide crucial constraints in the design of the several components used in the subsea. In this scenario we define **`Physical_Entity`** as *an entity that has physical existence*. Physical entities could be further subdivided in `Physical_Objects` and `Substances`. A **`Physical_Quantity`** is a physical property of a `Physical entity`, that can be quantified by a `Physical_measurement`. Sometimes `Physical_Quantities` can also refer to models of a `Physical_Entity`, so that they are also members of an Abstract entity (i.e. the model itself): the model may exhibit properties that are measurable through `Physical_Quantities.`

As discussed in the next sections, we are able to associate each text to one of these concepts. Being able to associate a sentence like "*The XMT is designed for 800 bar*" with the `Physical_Quantity` is a physical property of a `Physical_entity` i.e. `Pressure`, to a `Physical_quantity`, i.e. 800 bar, and to the `Type_of_Activity` that is `X-mas_trees_deployment`, we will be able to enable the query:

"Give me all requirements regarding pressure of the XMT."

by retrieving all semantic metadata that have been associated with the texts from the Requirment documentation.

In the next section a data driven method to associate texts to such Ontological concepts, is described.

Machine learning for semantic annotation

Statistical learning methods make the assumption that lexical or grammatical observations are useful hints for modeling different semantic inferences. Linguistic observations provide features for a learning method that are generalized into predictive components in the final model, induced from the training examples. In (Mitchell, 1997), Tom Mitchell provided an interesting definition of a learning program:

A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**.

In Natural Language Processing (NLP) such formulation allows to define learning systems that can be applied to Software Engineering. In particular:

- **T** represents a linguistic task, usually an interpretation, such as in semantic annotation or document classification tasks. In this study, such semantic processing task will be formulated as a Statistical Classification problem: the target is to identify the sub-population to which new data belong, where the identity of the sub-population is unknown (the test data), on the basis of a training set of data containing observations whose sub-population is known (the training data). For example, in the Document Classification task, texts are mapped to a set of classes that characterize the document topics, e.g. a document refers to sport, economics or politics. The objective is this the acquisition (from data) of a function $y = f(x)$ that is able to associate each text x to is corresponding class y.

- **P** represents the performances, thus measuring the quality of the resulting interpretation power. It depends on the task objectives and the learning system requirements. For example, if one is interested in the quality of a document classification system, the accuracy score can be employed as the percentage of correctly classified texts. However, if the learning algorithm improves the performance according to other aspects, e.g. the time needed for classification or the resource requirements of the produced learning system, other performance measures can be employed.

- **E** is represented by data as observations available about the target task. The idea is that a learning system exploits such information in order to acquire competences to resolve the target problem; the more information are observed, the highest are the performances P to solve the task T. In the classification task, experience is provided by the document themselves that are examples x providing different aspects of the target problem in terms of linguistic observations, such as lexical, grammatical or syntactic information.

Different Machine Learning algorithms exist in order to exploit data evidences and acquire a model of the target task, as discussed in (Bishop, 2006). In the results, the Support Vector Machine (SVM) learning algorithm, discussed in (Vapnik V. N., 1998) and (Basili & Moschitti, 2005), will be employed as it provides an effectively learning paradigm to satisfy our objectives. SVMs can be thought of as methods for constructing classifiers with theoretical guarantees of good predictive performance in terms of the quality of classification on unseen data. The theoretical foundation of this method is given by statistical learning, discussed in (Vapnik V. N., 1998).

More formally, the goal of a statistical learning algorithm is to learn a mapping from inputs x to outputs y, where $y \in \{1, \dots, C\}$, with C being the number of classes. If $C = 2$, this is called **binary classification** (in which case we often assume $y \in \{0, 1\}$); if $C > 2$, this is called **multiclass classification**. One way to formalize the problem is as **function approximation**. We assume $y = f(\mathbf{x})$ for some unknown function f, and the goal of learning is to estimate the function f given a labeled training set, and then to make predictions using $\hat{y} = \widehat{f(\mathbf{x})}$, a function estimation. Our main goal is to make predictions on novel inputs, meaning ones that we have not seen before (this is called **generalization**), since predicting the response on the training set is trivial. So, we need data to construct prediction rules, often a lot of it. We thus suppose we have available a set of measurements (x_i, y_i) or (x_i, g_i), $i = 1, \dots, N$, known as the training data, with which to construct our prediction rule. As discussed later, in this study we will consider an ontology that model the Oil Extraction within the Subsea domain. In our modeling, each concept within this ontology represents a class. Sentences are linguistic objects modeled through artificial representation **x**. Example of binary classification is the recognition if a sentence represents or not a requirement. Examples of multiclass classification is the classification and assignment of sentences expressing requirements to concept of the ontology.

Robust learning with SVMs

Support Vector Machines (Vapnik V. N., 1998), (Basili & Moschitti, 2005) represent a classifier belonging to the family of kernel methods, based on statistical learning theory, that proved to outperform many other categorization algorithms (Wa'el Musa Hadi, 2007), being able to handle large vector spaces with excellent accuracy (Xue Li, 2005).

Classifying data is a common task in Machine Learning, and with respect to the System Engineering, it allows classifying requirement according to discrete score. Here requirements are sentences represented by points in a multi-dimensionality space that can be assigned to a class in compliancy with their type or capability.

In a supervised-learning perspective, a training set of data point is given, each of which belonging to one or more class class, and then, the goal is predicting the class a new data point from a disjoint testing set, will be assigned, on the base of what learnt from the given training set. Support Vector Machines will be the method employed in this project for the classification of requirements among given classes of ontological types or capabilities. First suggested by Vapnik in the 1960s, SVMs are closely related to neural networks and they perform classification by constructing an *m*-dimensional hyperplane that optimally separates data into two classes (classical formulation) as well as in multiple classes. SVMs rely on kernel functions and, depending of these, they are an alternative training method for polynomial, gaussian functions and multi-layer perceptron classifiers (like sigmoid) in which the weights of the network are found by solving a quadratic programming (QP) problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem. In a common binary classification task, SVMs look for the optimal separating hyper-plane between the two classes by maximizing the margin between the classes' closest points. Points lying on the boundaries are called *support vectors*, and the middle of the margin is the optimal separating hyper-plane.

The classification task has been formulated in (Vapnik V. N., 1998) as a quadratic optimization problem, which can be solved by known techniques, employing Lagrangians, and this is the problem solved by the SVMs. Further, it can easily be extended to *k*-class classification, for example by constructing *k* two-class classifiers as discussed in (Rifkin & Klautau, 2004). Thus, considering the linear separable case, being:

- X The set of training data points, for which classes are known,
- $x \in X$, a data point,
- w a weight vector, normal to hyperplane

Then following function defines a separating hyperplane:

$$\sum_{i=1}^{m} w_i x_i + b = w^T x + b = 0$$

And the decision to establish the class of a new point is given from the sign of such function. That is, the classifier, based on an inner product is:

$$f(x) = sign(w^T x + b) = \begin{cases} +1 \\ -1 \end{cases}$$

A linear classifier may not be the most suitable hypothesis for the two classes. When a linear separator cannot be found, data points are projected into a higher-dimensional space where the data points become linearly separable. Such projection can be realized through the so called ***Kernel Tricks***, and the inner product in the classifier function can be computed in the new space without an explicit mapping, avoiding time and memory consuming due to the high space dimensionality of new space. In the classification phase the hyper-plane is not directly defined, as it is the linear combination of SVs, but the classification is still feasible in terms of the similarity (the dot-product) among the novel instances and the support vectors. The explicit representation of the novel feature space is thus never built and it is thus called implicit feature space. Kernel Methods, e.g. (Shawe-Taylor and Cristianini, 2004), refer to a large class of learning algorithms based on inner product vector spaces, among which SVMs are one of the most well known learning algorithms. The learning algorithm will select the most representative instances and characteristics in the implicit space, i.e. the space dimensions. Such methods provide effective statistical predictions defining meaningful similarity (i.e. kernels) functions among examples. It allows defining a learning paradigm whereas the algorithm can be directly applied over complex linguistic structure. These effective models can be acquired without focusing the attention over artificial representations and a novel and more effective representation can be defined in the implicit projection space.

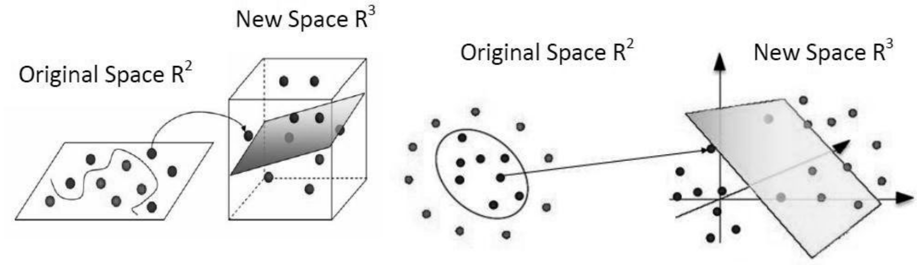


Figure 1 Examples of kernel trick mapping

Often flexibility faces with the difficult of modeling data with the right quantitative of inherent noise, providing a more robust generalization. Kernel methods are also beneficial because the combination of kernel functions can be integrated into the SVM, as these are still kernels, as discussed in (Shawe-Taylor and Cristianini, 2004). Consequently, the kernel combination $\alpha K_1 + \beta K_2 + \gamma K_3$ linearly combines in a disjunctive form the contribution of kernel K_1, K_2 and K_3. Here, parameters α, β and γ weight the combination of the three kernels.

SVMs are a powerful tool for binary classification, capable of generating very fast classifier functions following a training period. There are several approaches to adopting SVMs to classification problems with more than 2 classes (Rifkin & Klautau, 2004). The most common approaches are those that combine several binary classifiers and use a voting technique to make the final classification decision. A more complex approach is one that attempts to build one Support Vector Machine that separates all classes at the same time (Habib, 2008).

One-Against-All (Rifkin & Klautau, 2004) is the earliest and simplest multi-class SVM, as well as the approach adopted by the employed framework SVM-Multiclass. For a K-class problem, it constructs K binary SVMs. The *i*-th SVM is trained with all the samples from the *i*-th class against all the samples from the other classes. To classify x, the point is evaluated by all of the K SVMs and the class with the largest value of the decision function is selected.

Such an approach grants every point to be assigned with a certain class.

Language Technologies, Text Semantics and Classification

In the previous section we have described the statistical learning theory and, in particular, we have analyzed an efficient algorithm, Support Vector Machine. This algorithm learns from annotated data that is represented in a vector space. Because in many phases of the system life cycle, the documents are in natural language, i.e. technical documents, it is necessary to interpret a text to transform it into a vector. Representing the meanings of words and sentences in a form suitable for use by a computer is a central problem in Computational Linguistics (Jurafsky, 2000). The problem is of theoretical interest – to linguists, philosophers and computer scientists – but also has practical implications. Finding a suitable meaning representation can greatly improve the effectiveness of a Natural Language Processing (NLP) system. There have been two distinct approaches to the representation of meaning in NLP. The first, the *symbolic* approach, follows the tradition of Montague in using logic to express the meanings of sentences (Dowty, Wall, & and Peters, 1981). The logical representation of a sentence is built up compositionally by combining the meanings of its constituent parts. In contrast, the *distributional* approach uses statistics about the contexts in which a word is found, extracted from a large text corpus, to provide a vector-based representation of the meaning of an individual word.

The term *distributional* can be interpreted as *context-theoretic* since it encloses a rich family of approaches to semantics that share a usage-based perspective on meaning, i.e., they assume statistical distribution of terms in context plays a key role in characterizing their semantic behavior. Distributional models rest on the hypothesis that the degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear. Therefore, the more the linguistic context is representative of the distributional behavior of a term, the more semantic properties can outcrop inspecting it. Thus, distributional models are inherently related to statistics theory as well as to vector algebra.

With regard to the System Engineering, and then to the Requirement Analysis, one could imagine that in an evaluative text, term distribution may affect somehow the sentiment expressed. Thus, an interesting question would be if vector based models, representing statistical characteristics of texts are equally capable to catch the requirement semantics.

In this work we show here how Natural Language Processing and Machine Learning techniques are used to automatize semantic annotation of requirement documents by locating sentences expressing requirements and assigning them specific (ontological) types. Semantic Types refer here to the underlying Ontological model.

A Semantic Search Engine for Requirement Analysis

In this section, the architecture of the system is presented. Figure 1 shows it in order to emphasize the different involved modules.

- The **Import Handler** loads and pre-processes the requirement documents, in order to acquire a representation that is readable by the following modules.
- The **Versioning Handler** realizes the functionalities to handle the documentation versioning. When the upgrade of a given document is available, it retrieves differences between the documents in order to enable the tracking of the

requirements across upgrades. As an example, when a sentence expressing a requirement is added, the system will assign a proper type to the new sentence, and it will thus notify the addition of a novel requirement.

- The **Reveal Natural Language Toolkit (RevNLT[1])** implements techniques for natural language processing (NLP) in order to achieve a morphosyntactic analysis of the nature of the texts contained within documents. Example of this analysis is the segmentation of the documents into sentences, the identification of the main classes that characterize grammatically the words that make up the sentences (e.g. nouns, verbs or adjectives); parsing of sentences, which allows the extraction of linguistic constructs as the *Subject-Verb-Object Complement*. In the overall architecture, this system represents a module providing linguistic information useful to build artificial representation for the learning algorithm.
- The **Classifier Representation Builder** acquires each sentence, selects from the NLP system the proper information needed to generate the artificial representation useful for the learning algorithm. Some information can be derived by lexicon and terminology collection that represents *Domain Specific Semantic Metadata*. The lexical generalization acquired through the *Word Space* model, (Schütze, 1998).

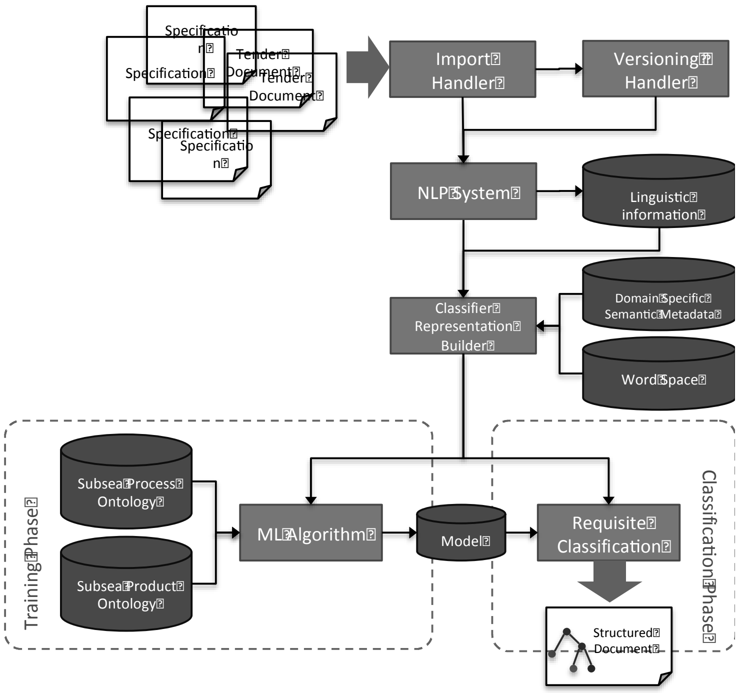


Figure 1: Overview of the system Architecture

Each representation follows a different workflow depending on the training or the classification phase. When acquiring the statistical model it is expected data to be labeled

[1] www.reveasrl.it

with *types* derived by the ontologies, i.e. **Subsea Process Ontologies** or **Subsea Product Ontologies**. The **Machine Learning Algorithm** loads labeled material in order to acquire a function to associate novel requirements and proper semantic types. This statistical ***model*** is then used by the **Requirement Classifier** to associate proper semantic type to novel sentences.

Experimental Evaluation

In this Section, the application of the proposed methodology to a real scenario and documentation derived within the Subsea domain is provided. The aim is to show the applicability of NLP and Machine Learning techniques in characterizing requirements and assign proper types reflecting the corresponding Work Package or Class Product; these classes are supposed to populate the Subsea domain ontology. In order to provide an exhaustive and quantitative evaluation, Aker has provided a dataset of requirements, labeled according to the targeted classes.

In the rest of this Section the evaluation metrics are discussed **.** The modeling used within the Classifier representation builder is discussed to present and discuss the evaluation of the acquired classifier with respect to the Work Packages and Product classes, respectively.

Performance evaluation in automatic classification system is needed to establish if the model acquired from the examples is suitable for use on future cases (usually called test cases) so that in the operational scenario the system is indistinguishable from the ground truth and the classification function has been perfectly learned.

The observation of the outcomes of a classifier *h(x)* allow to compute the following scores:

- *Error rate*, that is the ratio between the number of test cases not correctly classified (whereas the system prediction is not in agreement with the gold standard) and the total number of test examples processed. Dually, the percentage of the correctly classified examples constitutes the so-called *accuracy*;
- *Precision* e *Recall*, that better account for the system tendency to introduce mistakes (false positives) or miss correct decisions (i.e. false negatives), depend on a category C_i (for example, the *i*-th emitter E_i in the library), and the following quantities[2]:
- the number *tp* of *true positives*, i.e. correctly classified examples, that is the number of examples $x \in T$ such that $C_i \in h(x)$ and $C_i \in GS(x)$;
- the number *fp* of *false positives*, that is the number of examples such that $x \in T$ and also $C_i \in h(x)$ with $C_i \notin GS(x)$;
- The number *fn* of *false negatives*, that is the number of examples such that $x \in$ with and $\in$).

Given the above definitions also the following definitions hold:

[2] The name GS represents here the *Gold Standard,* that is the collection of annotated examples *x* labeled in the different classes C_i (when $C_i \in GS(x)$) or outside these classes (i.e. $C_i \notin GS(x)$).

In general, *Precision* and *Recall* vary in an inversely proportional manner. Moreover, as a single measure is often preferable, a largely adopted combination function is the harmonic mean of precision and recall that gives rise to the following score called F1 measure:

Let us consider the sentence "*The flow line shall therefore be controlled by the Control System XY123*". The correct association of this sentence with the class `Control system` is considered a *true positive*. On the contrary an incorrect association will generate a *false negative,* as the class has not been recognized. Moreover, if the incorrect association relates the sentence with the class `Umbilical`, it will be considered a *false positive* for this last class.

Modeling requirements within kernel based learning

The *Recognition Requirements Relation System* has been employed to recognize derive relation between requirements. The module employs Support Vector Machine classifiers to recognize correct pair of requirement. For different approaches are employed to model pairs, in order to investigate different linguistic information suitable for the target task and identify the most informative representation for the learning algorithm.

Let us consider the requirement "Self tests of the XMT shall be implemented", we extract four different representation

- The **full-text *Bag-of-Word*** (ft-*BoW*) model accounts mainly on lexical information, resulting in a vector whose dimension reflects the different words composing each requirement. As an example, dimension are "self" "*test*" or "*XMT*";
- The **linguistic processed *Bag-of-Word*** (lp-BoW) model: each sentence is processed by the NLP system; it allows to provides more expressive information to the learning algorithm, exploiting external and domain specific semantic information. As an example dimension are "*self test*", that carries out a more specific sense of "*test*", ore "*Christmas tree*" that is the meaning of the acronym XMT. This will extend the simpler information provided by the full-text Bag-of-Word; the linguistic analysis allow also to normalize information, avoiding data sparseness: as an example the verb "*implemented*" is replaced by the infinite version "*implement*"
- The **entity-driven Bag-of-Word** (ed-BoW) model: it express the linguistic filtered by the word part-of-speech. It focus more on the objects or events, then filtering out only nouns;
- The **Latent Semantic Analysis** (*LSA*) model accounts on lexical information enriched with expansion distributional models for represents meaning of words. The Word Space is derived through the distributional analysis of a document collection made available from Words are projected into a geometrical space derived through the distributional analysis introduced earlier in this paper. Each word is represented as a vector and then linearly combined in order to have a synthetic representation of the entire sentence.

Each representation enables the application of a different kernel functions. The first BoW models are used in linear kernels, while the LSA is used in a Gaussian kernel. The combination of kernel functions is applied: consequently, the kernel combination

$$\alpha \cdot ft\text{-}BoW(x,y) + \beta \cdot lp\text{-}BoW(x,y) + \gamma \cdot ed\text{-}BoW(x,y) + \delta \cdot LSA(x,y)$$

linearly combines lexical properties captured by BOW kernel and generalized by LSK. Here, parameters α, β, γ and δ weight the combination of the three kernels.

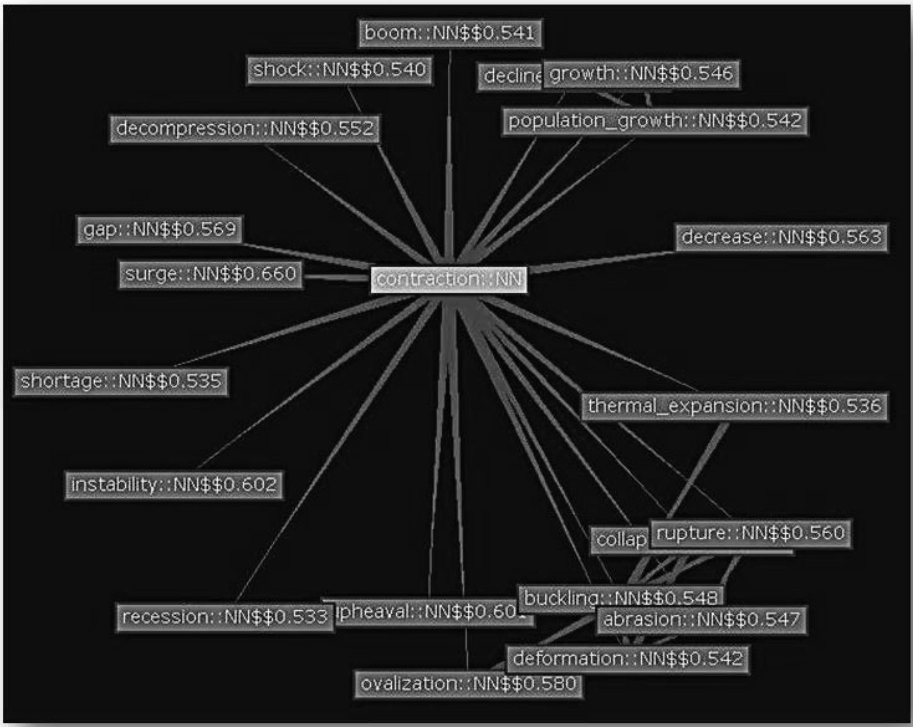
Early application: recognizing correlated concepts

In an early test of the concept, one query of choice has been a noun, namely “contraction”.

Among the other terms semantically close to “contraction”, the system returned “upheaval buckling”.

This term has a very specific connotation in the subsea pipeline industry, describing the phenomenon where a pipeline, partially or totally trenched and thus impeded in its free movement, exhibits a sharp bend protruding from the soil in a location that gives way to the pipe that the hot flowing oil has thermally elongated.

This connection is not at all trivial and only a minority of practitioners in the subsea industry is familiar with the term “upheaval buckling”. Of relevance, is the clustering of highly correlated concepts like ovalization, deformation, collapse.



2-d plot of the semantic query "*contraction*". The distance among the text boxes in the graph is proportional with the semantic distance among the terms contained by each box

Case Study: Recognizing Activity Types

This use case presents and discusses the evaluation of the acquired classifier with respect to the **Activity Types**, as defined in Aker Solutions, i.e. Work Packages (WP) classes. The aim is to acquire a classifier, i.e. a function, able to associate each sentence expressing a requirement to a specific WP.

The Support Vector Machine learning algorithm has been employed. As the association with WP classes is a multi-classification task, several binary SVM are trained according to the One-VS-All schema, as discussed in (Rifkin & Klautau, 2004). Each novel requirement is processed by all binary classifiers, each one representing a possible WP. It means that, when a classifier provides a positive response, the corresponding class is assigned to the text. When no classifier expresses a response, the text is not recognized as a requirement.

Labeled sentences from a training corpus provided by Aker Solutions have been considered. In order to avoid introducing entropy in the learning phase, caused by requirement assigned with too many classes, all training examples assigned with more than three classes are ignored. It is reasonable as they reflect too generic examples that are not useful to discriminate single classes.

Table 1 and Figure 2 show the number and the distribution of training examples for each class. Notice that classes with less than 20 training examples are considered as not sufficiently represented by training material and neglected. Classifier parameters are tuned with a Repeated Random Sub-sampling Validation, consisting in a *10-fold* validation strategy on a subset of the training data split according to a 90%-10% proportion. Table 2 shows the resulting number of training example. As each requirement can be labeled with more that one class, the number of single requirements is less than the number of possible (*requirement*,*class*) or (*r*,*c*) pair.

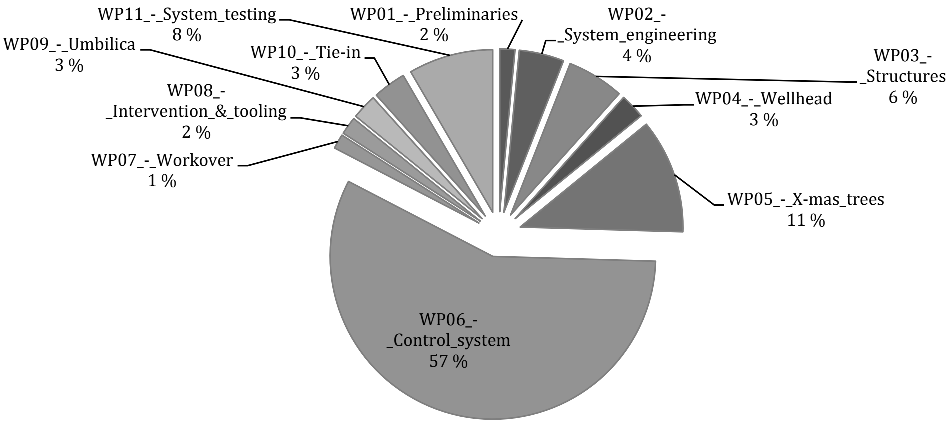


Figure 2 WP dataset: data distribution

Activity Type, WP class	Number of (*r*,*c*) pairs
WP01 - Preliminaries	29
WP02 - System engineering	86
WP03 - Structures	10
WP04 - Wellhead	48
WP05 - X-mas trees	21
WP06 - Control system	102
WP07 - Workover	28
WP08 - Intervention & tooling	34
WP09 - Umbilical	48
WP10 - Tie-in	63
WP11 - System testing	16

Table 1 WP dataset: number of training example for each class

Number of requirements	1412
Number of (requirement, class) pairs	1928
Number of classes	11
F1 Baseline Random	19,16%

Table 2 WP dataset statistics

Results are shown in Table 3, where the evaluation metrics have been applied. Mean results among 10 different runs show that the 87,68% classes, i.e. semantic types explaining requirements, proposed by the system are corrected. On the other hand the 79,41% of labeled assigned by the annotator have been recalled by the system. This good tradeoff between discovered information and well founded responses are confirmed by the F1 of 83,29%. These are promising results, especially compared with the baseline: in fact, by assigning a random class to different requirements, the system achieves a F1 of 19,16%, as shown in Table 2.

	TP	FP	FN	precision	recall	f1
Run1	149	18	46	89,22%	76,41%	82,32%
Run2	141	14	46	90,97%	75,40%	82,46%
Run3	145	24	45	85,80%	76,32%	80,78%
Run4	161	13	31	92,53%	83,85%	87,98%
Run5	151	34	53	81,62%	74,02%	77,63%
Run6	152	23	37	86,86%	80,42%	83,52%
Run7	159	18	37	89,83%	81,12%	85,25%
Run8	152	29	30	83,98%	83,52%	83,75%
Run9	155	21	32	88,07%	82,89%	85,40%
Run10	153	21	38	87,93%	80,10%	83,84%
Mean	151,8	21,5	39,5	87,68%	79,41%	83,29%
St.Dev	6,0	6,5	7,7	3,27%	3,60%	2,81%

Table 3: WP dataset experimental results. A Repeated Random Sub-sampling Validation, consisting in a *10-fold* validation strategy on a subset of the training data split according to a 90%-10% proportion has been applied

Conclusions

Although the adopted models and the experiments carried out in the investigation described in the above sections are not to be considered definitive, a set of early outcomes has been derived and allow us to draw some consequences. Ontological Knowledge as implicit in the analyst work on text documentation is a building block for the automation of a number of functionalities, such as *locating* the proper information (e.g. reference to ISO standard documents) or *capturing the relationship* between a requirement and the device is responsible to satisfy it, useful to easy the analyst's task and improve its productivity. The explicit representation of ontological notions such as all domain concepts, relations and properties, has a strong impact on the accuracy and costs on the engineering of any intelligent software platform for supporting Requirement Analysis. Extracting conceptually relevant information from the text involved in RA is complex, but it can be cost-effectively automatized by the help of inductive technologies such as SVMs. These enable a direct application of the Ontology model as they are able to automatize the recognition of some of the semantic properties declared in the Ontology within an incoming text (e.g. a project design document as well as a tender), so that new texts are harmonized with previous information/knowledge whatever its source is. The relatively short time required to develop a semantic annotation tool for documents entering in the RA chain is a clear evidence of the potentials of these method in realistic operational RA settings. In synthesis, we showed that Natural Language Processing and Machine Learning can be effectively used to enable automatic semantic annotation of requirement documents by locating sentences expressing requisites and assigning them specific (ontological) types. Semantic Types used here refer to a basic Ontological model and are based on strict adoption of WWW standards. The quantitative benchmarking reported confirms the viability of the approach.

References

ISO 13628-1 Petroleum and natural gas industries -- Design and operation of subsea production systems -- Part 1: General requirements and recommendations

Basili, R., & Moschitti, A. (2005). Automatic Text Categorization: from Information Retrieval to Support Vector Learning. Aracne Ed.

Christopher M. Bishop. 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Dowty, D. R.; Wall, R. E. & Peters, S. (1989), Introduction to Montague Semantics , Vol. 11 , D. Reidel , Dordrecht .

Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43, 5-6, 907-928.

Habib, M. S. (2008). Addressing Scalability Issues of Named Entity Recognition Using Multi-Class Support. World Academy of Science, Engineering and Technology, International Science Index 13, 2(1), 61-70.

Daniel Jurafsky and James H. Martin. 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (1st ed.). Prentice Hall PTR, Upper Saddle River, NJ, USA.

Maedche, A. & Staab, S. (2000). Discovering conceptual relations from text. In Proceedings of ECAI-2000. IOS Press, Amsterdam.

Thomas M. Mitchell. 1997. Machine Learning (1 ed.). McGraw-Hill, Inc., New York, NY, USA.

Rifkin, R., & Klautau, A. (2004). In Defense of One-Vs-All Classification. The Journal of Machine Learning Research , 5, 101-141.

Hinrich Schütze. 1998. Automatic word sense discrimination. Journal of Computational Linguistics. 24, 1 (March 1998), 97-123.

Shawe-Taylor, J., Cristianini N. 2004. Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

Vapnik, V. N. (1998), Statistical Learning Theory , Wiley-Interscience.

Hadi, W. M.; Thabtah, F. A. & Abdel-jaber, H. (2007), A Comparative Study using Vector Space Model with K-Nearest Neighbor on Text Categorization Data., in Sio Iong Ao; Leonid Gelman; David W. L. Hukins; Andrew Hunter & A. M. Korsunsky, ed., 'World Congress on Engineering' , Newswood Limited, pp. 296-300.

Xue Li, Shuliang Wang, Zhao Yang Dong (Eds.): Advanced Data Mining and Applications, First International Conference, ADMA 2005, Wuhan, China, July 22-24, 2005, Proceedings. Lecture Notes in Computer Science 3584 Springer 2005, ISBN 3-540-27894-X.

Biography

Luca Abele Piciaccia developed his 25+ years systems engineering career in the Subsea Oil Industry at major oil companies and leading EPC contractors, where he held several positions as Engineering Manager and Commissioning Responsible before becoming Chief Engineer for Subsea Systems Engineering - Major Subsea Projects at Aker Solutions. His education includes Nuclear Engineering at the Technical University in Milano, BSc in Engineering Management at CCU and an MBA in Technology Management from Chiefly Business School in Australia. Luca has served 2 terms as president of the Norwegian INCOSE chapter.

Roberto Basili is Associate Professor of Computer Science at the University of Roma, Tor Vergata, where he carries out research and development on Natural Language Processing, Machine Learning, Information Extraction and Retrieval. Currently, he is the coordinator of the SAG@ART group and technical responsible of the AI Laboratory of the Enterprise Engineering Department. His research interests include Representation and Learning of Lexical Knowledge, Information Retrieval and Extraction, as well as Ontology Engineering. He acted as unit technical coordinator within several European Projects (NOMOS, TREVI, ECRAN, NAMIC, PrestoSpace, INSEARCH) on advanced applications of language technologies to human computer interfaces or Web retrieval services. He is author of more than 100 papers on international journals, books and conference proceedings.

Danilo Croce has Ph.D. in Informatics Engineering from the University of Roma "Tor Vergata". Currently, he is a postdoc and he is a member of the SAG@ART group at the Enterprise Engineering Department in the same university. His expertise concerns theoretical and applied Machine Learning in the areas of Natural Language Processing, Information Retrieval and Data Mining. In particular, he is interested in innovative kernels within support vector and other kernel-based machines for advanced syntactic/semantic processing.

Valerio Storch holds a Master Degree in Computer Engineering in 2011, from the University of Roma Tor Vergata. He is software analyst at Reveal srl and participates to the the SAG@ART group at the university. His main research area are Semantic Search and Information Retrieval.

24th Annual INCOSE International Symposium (IS2014)
Las Vegas, NV, June 30 – July 3, 2014

Requirements Elicitation through Semantically Aware Techniques for the Subsea Oil Industry Systems Engineering

Luca A. Piciaccia
Aker Solutions
Luca.Piciaccia@akersolutions.com

Roberto Basili
University of Roma, Tor Vergata
basili@info.uniroma2.it

Danilo Croce Ph.D.
University of Roma, Tor Vergata
croce@info.uniroma2.it

Cecilia Haskins, ESEP
NTNU, IPK
cecilia.haskins@ntnu.no

Abstract. Based on an innovative application of systems engineering in the subsea oil industry, the text illustrates how Semantically Aware Retrieval and Browsing (SARB) and Simple Knowledge Organization Schemes (SKOS) have been successfully used for the first time to efficiently identify and make available concepts and requirements. The techniques focus on eliciting the semantic connections leading to implicit requirements. In addition, the methodology has been successfully applied to real time lessons learned semantic mining. The study will be expanded and applied to frontier research in requirement management and design review. The method is currently being applied for a six months trial in five major subsea oil projects to verify its performance in an industrial environment.

Introduction

The framework of this study is the systems engineering activities in the subsea oil industry, specifically on the challenges of the requirement elicitation phase. In the subsea oil industry the relevant requirements are often embedded through implicit concepts in the documentation or addressed by references to the current normative bodies.

The systems engineering mission in the subsea industry can be summarized as follows (adapted from the IEEE definition):

- Translate an operational need into a configured system
- Integrate all technical disciplines in a coordinated effort that meets cost schedule and performance
- Ensure compatibility of interfaces
- Ensure design meets requirements
- Measure and control technical risk

Background

Requirement elicitation is a challenge in the subsea oil extraction industry due to the short project award and execution times that do not permit a complete compilation of a requirement database from the contract documents and appendixes. This results in a risk based approach to counter the adverse effect of the incomplete information and the inevitable late discovery of some requirements with *ad-hoc* mitigations. This study is the first step in research toward an improved method to elicit requirements in the subsea oil industry. The objective of the research is to open a new path in requirements analysis and deliver an enabling technology for mainstream requirements management techniques using COTS (Commercial-Off-The-Shelf) products.

Research criteria. One success criteria of this research is to provide an intelligent requirement elicitation environment that improves the full compilation of the requirements register with the necessary quality and within the short project award cycles typical of the subsea oil industry. The current state of the art relies on lengthy documents compiled in natural language and purporting numerous references to additional texts and appendixes to describe the scope of work for the contract being awarded. This results in an involute description where lack of clarity is rife and the very nature of natural language inhibits an efficient individuation of the requirements.

The fulfillment of a requirement is achieved when said requirement is tested and the test results meet the agreed acceptance criteria. It is self-evident that poor requirement clarity will inevitably lead to a difficult test regime and increased risk. This research intends to prove that it is possible to effectively support the expert analyst in the task of requirement retrieval whilst navigating huge document collections dealing with requirement engineering in the different stages of the process (e.g. tender analysis, system design or system validation).

Research questions. The fundamental question for this research is as follows: is it possible to build and put to use an ontology or at least a taxonomy for the subsea industry and should it be built top-down (axiomatic) or bottom-up (semantically)?

Semantically Aware Retrieval and Browsing (SARB). This question reflects a vision where a Semantically Aware Retrieval and Browsing (SARB) system is made available for the expert, able to embody his/her own view of the domain and its specific expertise. This in turn raises a number of questions regarding representation and user involvement.

Representation. How can we automate the inference needed by the system for realizing the semantic awareness? How can we automatically acquire the required Semantic Representation? How can we use the available corpora to make this acquisition cost-effective?

The options are to model the domain in terms of axioms from which the correct retrieval choices can be *deduced* (ontology as domain axiomatization) or, alternatively, to make the retrieval process semantically aware by making it consistent with conceptual aspects (for example technical terminology, domain specific similarity metrics and relevance criteria) of the domain using document classes (i.e. ontology as a document concept scheme, e.g. SKOS: Simple Knowledge Organization Scheme). The Axiomatic approach has several drawbacks described in the 'Methods' section of this text, therefore, the approach of choice for the research has been SARB acquisition based on SKOS.

User involvement. How can we make the user pro-active in the individual stages of the resulting workflow; namely, setting the lexical representation (e.g. validating terminology); setting the concept scheme (e.g. defining novel class labels or validating document clusters suggested by the system); interactive training, i.e. validating system outcomes in a continuous life-cycle, in order to make the system semantics as much harmonic as possible with the expectations of the analyst.

The complete the set of research questions can then be laid out follows:

1. Is it possible to build Ontology (or at least Taxonomy) for the subsea oil industry able to guide and optimize the general retrieval and browsing tasks?
2. How can we acquire and validate the above by mining the document sources in an automatic or semi-automatic fashion?
3. Which are the criteria to adopt in order to measure the quality and completeness of the Ontology within its practical uses in the industrial environment?
4. How can we design an intuitive and efficient man-machine interface able to optimize the user involvement in the retrieval and browsing processes: validating source information, performing individual queries, refining them and validating the corresponding results?
5. How can we design the proper integration of ontological knowledge (i.e. information described in knowledge representation languages, such as description logics) and textual evidences (e.g. lexical and statistical information) in order to make the resulting SARB system maximally beneficial to the overall project execution effort, in the industrial setting?

Test results for first phase of research

The first phase of the research tested the hypothesis that it is indeed possible to extract a practical ontology and put it to use in the subsea oil industry to augment the quality and efficiency of the engineering work and contribute to a significant reduction in cycle time for the engineering analysis of a project. The logical process can be described as applying domain ontology to the assertion documents to elicit the valid associations between the document at hand and the specific knowledge of the domain possessed by the experienced engineer. These associations, weighted by their corresponding "semantic distance", are presented as text extracts from the document itself based upon a query on one or several topics of interest.

Example: in an early test of the concept, one query of choice has been a noun, namely "*contraction*". The system returned among the items with the shortest semantic distance its antonym: "*expansion*". This is quite a simple and obvious connection, which is general and might even be non-domain dependent. Figure 1 shows a 2-d plot of the semantic query "contraction" in which the distance among the text boxes in the graph is proportional with the semantic distance among the terms contained by each box.

Among the other terms semantically close to "*contraction*", the system returned "*upheaval buckling*". This term has a very specific connotation in the subsea pipeline industry, describing the phenomenon where a pipeline, partially or totally trenched and thus impeded in its free movement, exhibits a sharp bend protruding from the soil in a location that gives way to the pipe that the hot flowing oil has thermally elongated. The pipeline exhibits the localized buckling behavior typical of its being a slender structure (diameter/length ratio > 20). This connection is not at all trivial and only a few practitioners in the subsea industry are familiar with the term "*upheaval buckling*". Of relevance, is the clustering of highly correlated concepts like *ovalization*, *deformation*, *collapse*.

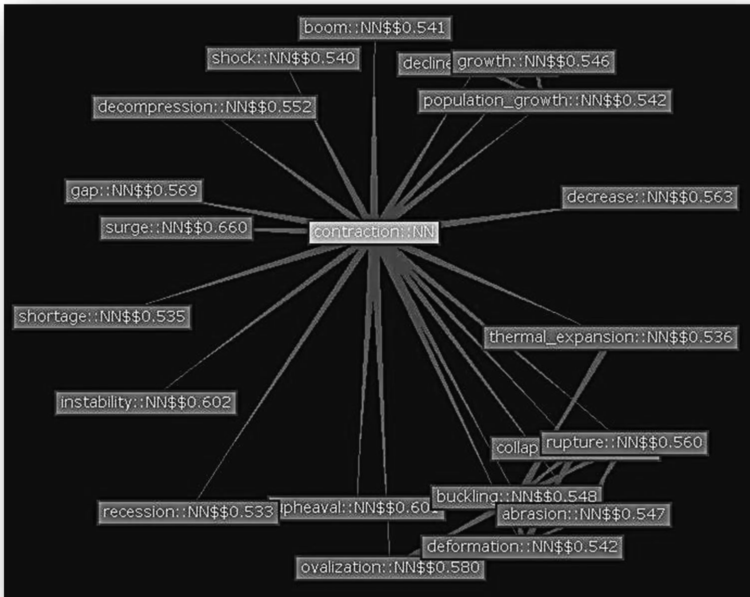


Figure 1. A 2-d plot of the semantic query “contraction” in which the distance among the text boxes in the graph is proportional with the semantic distance among the terms contained by each box

Support Vector Machine (SVM) learning algorithms, introduced in (Vapnik 1998) and more recently employed in text semantics (e.g. in (Basili & Moschitti 2005)), have been employed for this test, as it provides an effectively learning paradigm to satisfy our objectives. SVMs can be thought of as methods for constructing classifiers with theoretical guarantees of optima predictive performance in terms of classification accuracy on unseen data. The theoretical foundation of this method is given by the statistical learning theory, discussed in Vapnik (1998).

Semantic kernel functions for SVM learning can be used as models of text semantics. This perspective, together with numerous similar ones, corroborates the idea that the method is indeed capable of correctly identifying semantic similarities and correlations between the domain and the textual phenomena (such as assertions), thus conveying results of a quality similar to an analysis performed by a trained practitioner.

Method

The research approach has followed the general guidelines sketched by Eisenhardt (1989) and integrated by Valerdi and Davidz (2009) for empirical research in systems engineering.

Selecting Cases. The study was started as an emergent technology effort sparked by the presentation of AI assisted Ontology work at the 2012 INCOSE Symposium by Prof. Roberto Basili from the University in Roma Tor Vergata. Particularly striking was the approach to concept correlation, achieved by the system through "semantic distance", as this was mimicking the most successful approach in design reviews. Prof. Basili arranged for the initial early-phase/feasibility study as a pro bono effort. After this successful phase, a contract was signed between the University and Aker Subsea in 2013 for a development project.

Foundation for the research. The approach of inferring relations by applying domain knowledge to a natural language text is in itself instinctively appealing. After all, it is what we do when we submit a document to an expert to have it explained to us or commented. We expect the expert in the domain to point us to the meaning of each section and the connections it has in relation with the overall text and the domain in general.

This approach, although well ingrained in our ways as human beings, is subject to at least two major flaws, i.e. limitations in quality and completeness. The expert shall have all the available domain knowledge at his disposal (quality) and he shall be in a position to examine the whole text (completeness) to return the correct answer we are looking for within a useful response time. In real life, assuming just the domain knowledge collation of applicable industry standards listed in table 1, the expert(s) should be able to effectively retrieve and use information scattered across approx. 10,000 pages. This information should then be applied to the assertion document, in this case the contract and its appendixes, which itself is in the order of 1000 pages already.

It is self-evident that this approach is unmanageable in a fully deterministic way unless the cycle time of the approval process is measured in years. Beyond being practically awkward, this approach is also based on a logical foundation that can and should be questioned, i.e. the knowledge on a domain is an a priori determined framework whose taxonomy is well defined before any use of said knowledge can be made. The approach used in this work is a clear departure from such current praxis.

The main components of a domain knowledge model are well described in the Description Logic Handbook, as follows: (Nardi and Brachman 2003:13)

"Within a knowledge base one can see a clear distinction between intensional knowledge, or general knowledge about the problem domain, and extensional knowledge, which is specific to a particular problem. A typical Description Logic knowledge base analogously comprises two components – a TBox and an ABox. The TBox contains intensional knowledge in the form of a terminology (hence the term "TBox", but "taxonomy" could be used as well) and is built through declarations that describe general properties of concepts."

"The ABox contains extensional knowledge – also called assertional knowledge (hence the term "ABox") – knowledge that is specific to the individuals of the domain of discourse. Intensional knowledge is usually thought not to change so much –to be "timeless", in a way – and extensional knowledge is usually thought to be contingent, or dependent on a single set of circumstances, and therefore subject to occasional or even constant change." These statements can be held valid for all knowledge based systems and have no link to any supporting media; they are even applicable to oral traditions and non-explicit knowledge.

In laymen's terms the T-Box is the knowledge an engineer has accumulated through years of experience whilst the A-Box is "the problem at hand", be it the conceptual information as embodied by a specific document or a design report. The knowledge has a longer evolution time and tends to be stable, being based on facts and hard experiences whilst the work onto which the knowledge is applied (A-Box) is ever changing. The "conscious knowledge" is also referred to as Ontology, in laymen's terms "what we know about a subject".

The departure from the traditional approach is exactly here: the practical ontology for subsea oil production systems has been developed in this work as a "side effect" of the text analysis. Traditionally the collective knowledge of the engineering pool would be collated in a manual or a set of procedures and specifications edited by a dedicated expert or panel and made available for consultation and experience transfer. All in natural language and mediated by the engineer applying the manual to the new text to spot and classify compliance or divergence. Again, in the traditional attitude, the focus is on the domain knowledge as embedded in the expert panel brain, not in the neutral collation of fundamental guiding documents about the domain.

What has been done differently here is on the use of natural language itself as the backbone of domain semantics, in a sense a sort of implicit T-Box. This allowed leveraging the availability of the extremely refined and reliable natural language processing methods and tools refined through years of world-wide research over a large number of domains. The unique advantage is clear; the documents that normally express facts and norms of interest can now be analyzed semantically. As their content can be fully indexed and semantically organized for the underlying AI system, explicit inferences can now be drawn. They thus implicitly embody domain knowledge made semantically active throughout a variety of textual inferences.

These inferences can be applied to the facts the engineer need to reason about, i.e. queries about requirements at hand on a use case, as against a portion of a A-Box, in order to retrieve involved concepts and rank them by semantic relevance. Notice how these inferences (made available as in an A-Box) will augment the engineer information about the domain (experience) and be made available in the future cases in a dynamic learning environment: this process allows capitalizing the overall accumulated experience over the processed material and accessing it in an automatic fashion.

In this way an overall Ontology is built by exploiting the actual use of expert's language, and is not a filtered and biased reconstruction of the knowledge made by hand. In addition, linguistic analysis of the texts is automated so that semantic inferences able to identify and verify consistency, duplications, syntax and atomicity of the statements or requirements are made available as a side-effect.

Sources. The first conscious choice has been the selection of the relevant material to be collated for the ontology. Leveraging 25 years of domain expertise and systems engineering in the subsea oil industry the selection process converged to the material itemized in Table 1. Once the material was collected, it was semantically processed by a corpus processor. This analysis delivered a 1600-element-long list of terminological concepts that were semantically valid but still lacked validation in the domain of choice.

Table 1. Selected documents

Document Type	Document description
ISO 13628 -1 to 11	This is the industry standard in term of Subsea Oil Production systems
API 6 and 17	The industry standard from the American Petroleum Institute
ISO/IEC 15288	Systems Engineering – Lifecycle Processes
	Standing Technical document from major Oil companies
	Proprietary Design Basis and Design Review procedures
	Full documentation of a relevant subsea oil project from tender to contract to execution phase, delivery and closeout report.
	Relevant selected material in public domain on the internet.

SKOS buildup. Once the sources have been fixed, it is important to provide a structure for the concepts they are dealing with. This is what we intend by organization scheme, designed through the SKOS standard. SKOS (Simple Knowledge Organization Scheme) is a classification standard used to represent term and document lists, controlled vocabularies and thesauri. The Technique enables direct Lexical labeling and supports simple broader/narrower hierarchies (with no formal semantics). SKOS enables a simple, machine-understandable, representation framework for Knowledge Organization Systems (KOS). The representation exhibits the flexibility and extensibility to cope with the variation found in KOS idioms. This is an essential feature when processing natural language. SKOS is fully capable of supporting the publication and use of KOS within a decentralized, distributed, information environment such as the worldwide (semantic) web. This feature is a necessity for contemporary IT architectures. SKOS provides a usable model for expressing basic structure of "concept schemes" and supports the necessary Thesauri, classification schemes, taxonomies and other controlled vocabularies, which allow the mining of the knowledge base with broad spectrum queries.

Concepts Identification and validation. In this phase, the list of proposed concepts had to be reviewed by domain experts to confirm or exclude each individual proposed concept as a valid one. The list of concepts is collated in concept schemes. A concept scheme is a set of concepts, potentially including statements about relationships between those concepts and augmented by the Semantic Relationships in their Broader/Narrower Terms. Related Terms are linked through Lexical Labels which are dynamically ranked in preferred, alternative and hidden labels. The ranking can be organized in such a way to learn from the choices made by users of the system and automatically adjusted to reflect semantic relationships that are discovered to be tighter or looser by their use in the domain of interest. The metadata associated with the concepts (origin and traceability) are fully preserved and can be augmented by additional documentation, notes, comments and descriptions. Concept schemes are not formal ontologies in the way that, e.g. OWL (Web Ontology Language) ontologies are formal ontologies.

Lexicalizing further the concept schemes. As SKOS provides a collection of mapping properties that express relationships between concepts in different schemes we mapped each concepts to a set of domain terms, automatically derived from the corpus, as discussed in (Basili et al, 2012). These latter are essential for increasing the semantic quality of the document retrieval and classification according to the SKOS hierarchy. In this case the most used mapping property is the broad/narrow Match as followed by the close Match and then exact Match. This choice is motivated by the fact that indiscriminate uses of stricter properties such as owl:sameAs can lead to undesirable inferences.

Example of SKOS concept taxonomy. As an example, four real world physical parameters were chosen as presented in Table 2 to show how a small excerpt of the simple taxonomy of the domain of interest is organized.

Table 2. A practical simplified subset taxonomy of the subsea oil industry domain

FIELD DESCRIPTION	DESIGN PARAMETERS & PREREQUISITES	SUBSEA PRODUCTION SYSTEM	TECHNICAL RISK AND SAFETY MANAGEMENT
Field layout	Design Water Depth	Scope of Supply	RAMS
Coordinates	Design Life	Tie-In and Tooling	Safety System Requirements
Water depth	System Availability	Controls	Systems shutdown levels
Fluid Compon	Design Weight	Chemical distribution system	
Flow forecast	Installation	Electrical distribution system	
Sand Production	Dropped objects	Umbilical Function List	
H2S Content	Overtrawlability	Umbilical System	
CO2 Content	Design Pressures	XMT	
Seawater for Water Injection Specification	Wellhead Shut-in Pressure	Production	
Environmental Conditions	WAG System	Water and Gas Injection	
Soil Conditions	Gas Lift/Injection		
	Water Injection		
	Design Temperatures		
	Material Selection		
	Insulation and Hydrate management		
	ROV intervention design criteria		
	Tagging		
	Chemical distribution system		
	Methanol distribution		

Concepts validation. Over 95% of the terms automatically identified by the corpus analysis phase within the domain specific SKOS hierarchy, in a fully unsupervised fashion (i.e. without any doctoring or previous training), were valid in the domain of interest. The validity was ascertained by having the list reviewed by domain experts who validated each term as belonging to the domain and in current use. Given the automatic nature of the applied process, this was the first concrete outcome confirming the validity of the underlying research perspective.

This in itself has a high significance. Relating the quality of this result with the expected one from an equivalent effort of a trained engineer tasked with the same assignment, it is safe to affirm that the quality level is absolutely comparable. I.e. the ontology generated through the innovative method of language analysis is as good as one generated through the traditional method of domain knowledge analysis.

The new method has multiple advantages. Cycle time: the concept list was generated in only few hours of machine elaboration, versus a comparable effort measurable in thousands of hours for a trained engineer. The knowledge base can be expanded by adding the new texts to the existing data and re-running the semantic analysis. All the metadata (e.g. title and page of the document in which each concept is found) are available and linked, enabling deep dives in search of the concept source, if needed. A semantic distance ranking is now possible, uncovering the possibility to identify clusters of closely related concepts. Such clusters can be of help in risk management efforts to assess consequence propagation paths.

Results

The Ontology was applied to the company lessons learned database, the repository of learnings accumulated in past projects, to supply a test bed for usability and real life application of the technology in the relevant industrial environment. To date this phase has delivered a fully usable solution that is being currently field tested. This step demonstrated the practical usability of semantic analysis and description logic for identifying risks (programmatic and technical) in the subsea oil industrial domain.

The first phase was concluded with a successful answer to the first four research questions presented above. The Ontology has been built and validated, with a surprisingly good adherence of the SARB/SKOS results with domain experts' opinion. The use of the retrieval interface has been demonstrated and the multiple domain experts who reviewed the phase 1 results do agree on the potential the technology has for augmenting the analysis capability of the engineering team.

An example of query output via the GUI is presented in Figure 2 and 3, where the closest match to the query of interest is represented graphically at the center in a 2-dimensional graph that conveys the semantic distance from each result and among all the results as distance on the plane. This method is able to convey the semantic associations of the various terms returned and augments the awareness of the operator to classes of phenomena, since clustering of results on the plane indicate a cluster of results tightly connected to each other. The complete metadata set for each result is available by simply clicking on the result of interest, this action returns a window showing the sentence that triggered the result and returns all available metadata for collection in export baskets enabling register compilation.

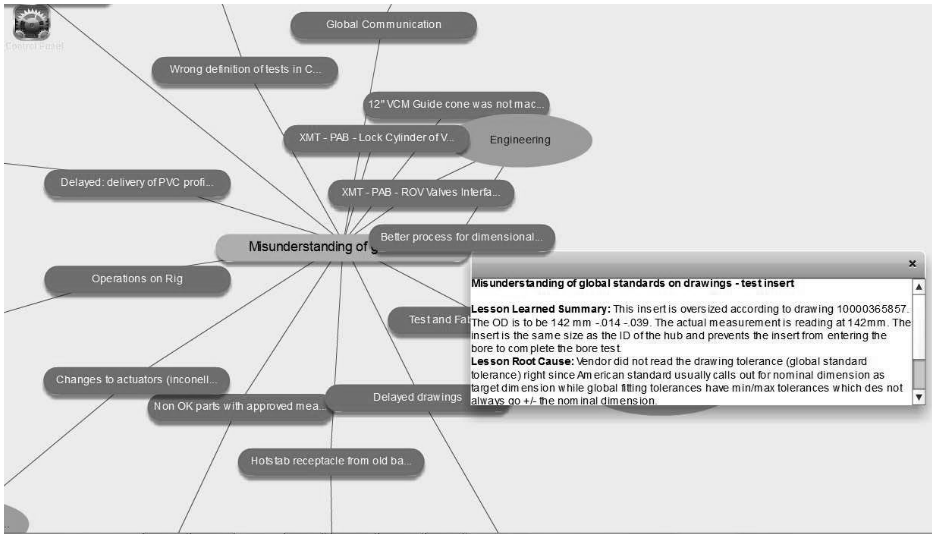


Figure 2: Semantic Graph: Example of one of the results on the query “standards” in the Lessons Learned A-Box.

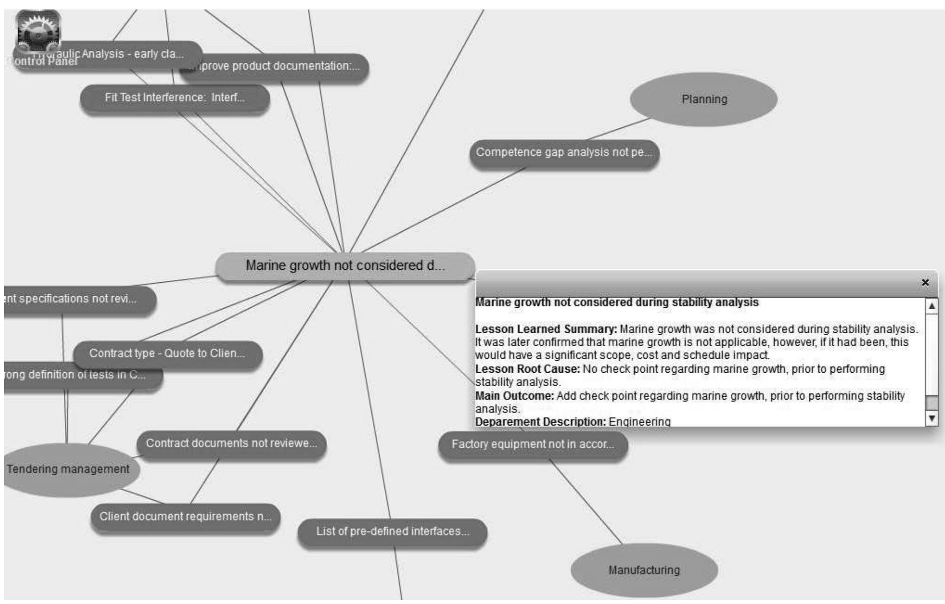


Figure 3: Semantic Graph: Example of one of the results on the query “marine_environment” in the Lessons Learned A-Box.

Conclusion

The work is of innovative nature for the oil industry and is considered, by the experienced practitioners consulted, to have large potential for improving the requirement elicitation quality and completeness. The initial application on the body of the lessons learned database has been an enabler for retrieval of "forgotten" requirements that at least in one occasion have crossed the line from "potential threat" to "actual threat", sometimes remaining undetected until the negative occurrence materialized in full.

Most important is the capability, enabled through SKOS, of retrieving concepts that are semantically linked without having to query on exactly the correct term. This has augmented the capability of the engineer to connect his search to many relevant domain topics (even those previously unknown to the individual) that are now presented for him/her to assess their relevance for the problem at hand.

The first Real time application during a major subsea project conceptual design review has enabled the discovery of one such threat, unknown to the project team, and the relevant mitigation action to be included in the design at an early stage. Although the adopted models and the experiments carried out in the investigation described in the above sections are not to be considered definitive, a set of early outcomes can be derived as follows:

Ontological Knowledge as implicit in the analyst work on text documentation is a building block for the automation a number of functionalities, such as locating the proper information (e.g. reference to ISO standard documents) or capturing the relationship between a requirement and the device is responsible to satisfy it, useful to easy the analyst's task and improve its productivity. The explicit representation of ontological notions such as all domain concepts, relations and properties, has a strong impact on the accuracy and costs on the engineering of any intelligent software platform for supporting Requirement Analysis.

The role of textual information is critical as it is source of ontological information (such as the terminology introduced by the ISO norms). It is also informative of the nature of a target project (e.g. the response to a tender), and it discusses its aims, constraints and requisites. Extracting conceptually relevant information from the text involved in Requirements Analysis is a complex task, but it can be automatized by the help of inductive technologies such as SVMs.

SVMs enable a direct application of the Ontology model as they are able to automatize the recognition of some of the semantic properties declared in the Ontology within an incoming text (e.g. a project design document as well as a tender), so that new texts are harmonized with previous information/knowledge whatever its source is. Moreover, all concepts recognized or assigned to a text through classification provide a relevant bod of evidence (semantic metadata) about texts content, so that effective indexing and retrieval is enabled for different stages of the RA process.

Experimental results carried out on a medium sized annotated data sets (i.e. examples used to train our SVM models) show how the semantic annotation task can be automated with great accuracy at different levels of the RA analysis. Baselines performance measures (with lowest baselines achieved with respect to deeper levels of the ontological hierarchy, i.e. at a finer grain) are significantly improved in all our tests. The relatively short time required to develop a semantic annotation tool for documents entering in the Requirements Analysis chain is a clear sanction of the potentials of these method in realistic operational RA settings.

In synthesis, we showed that Natural Language Processing and Machine Learning can be used to enable automatic semantic annotation of requirement documents by locating sentences expressing Requirements and assigning them specific (ontological) types.

Perspectives for future research. The study will be applied to frontier research in requirements management and design review. The aim is the reduction of technical and programmatic risk in projects characterized by high technological content and complexity.

Further activities will use the lessons learned as queries to elicit semantically connected concepts or clusters of concepts to be validated by domain experts for the specific case. The expectations are of a faster and more accurate compilation of the risk register verified for relevance, consistency and syntax. Subsequently, the first attempt for automatic or heavily assisted requirement identification and validation will be carried out against a recent past executed contract. In this way, a comparison of the theoretical results and a real life project whose life cycle has been completed will be possible. The aim is the compilation of a requirements register verified for consistency, duplications, syntax and atomicity of the statements or requirements. The requirement register will be in a format compatible with commercially available requirements management tools.

References

Roberto Basili, Armando Stellato, Daniele Previtali, Paolo Salvatore, Jörg Wurzer: Innovation-Related Enterprise Semantic Search: The INSEARCH Experience. ICSC 2012: 194-201.

Roberto Basili, Marco Cammisa and Alessandro Moschitti, Effective use of wordnet semantics via kernel-based learning. In Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL 2005), Ann Arbor(MI), USA, 2005

Danilo Croce, Alessandro Moschitti, Roberto Basili: Structured Lexical Similarity via Convolution Kernels on Dependency Trees. EMNLP, Edinburgh (UK), 2011: 1034-1046.

Eisenhardt, K. M. (1989). Building theories from case study research. Academy of management review, 14(4), 532-550.

Francesco Garzoli, Danilo Croce, Manuela Nardini, Francesco Ciambra and Roberto Basili. Robust Requirements Analysis in Complex Systems through Machine Learning, Joint workshop on Intelligent Methods for Software System Engineering (JIMSE2012) Montpellier, France, 2012.

Cristina Giannone, Roberto Basili, Paolo Naggar, Alessandro Moschitti: Supervised semantic relation mining from linguistically noisy text documents. IJDAR 14(2): 213-228 (2011).

Nicola Guarino, Daniel Oberle, and Steffen Staab: What Is an Ontology?, in Staab, S., Studer, R. (eds.) (2004), Handbook on Ontologies, Berlin, New York: Springer

Haskins, C., ed. 2010. *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities.* Version 3.2. Revised by M. Krueger, D. Walden, and R. D. Hamelin. San Diego, CA (US): INCOSE.

Alessandro Moschitti, Daniele Pighin, Roberto Basili: Tree Kernels for Semantic Role Labeling. Computational Linguistics 34(2): 193-224 (2008)

Alessandro Moschitti, Silvia Quarteroni, Roberto Basili, Suresh Manandhar: Exploiting Syntactic and Shallow Semantic Kernels for Question Answer Classification. Proceedings of ACL Prague, 2007

Nardi, D. and Ronald J. Brachman. "An Introduction to Description Logics" in Baader, F. (Ed.). (2010). The description logic handbook: theory, implementation, and applications 2.ed. Cambridge University Press.

Staab, S., Studer, R. (eds.) (2004), Handbook on Ontologies, Berlin, New York, Springer.

Valerdi, R., & Davidz, H. L. (2009). Empirical research in systems engineering: challenges and opportunities of a new frontier. *Systems Engineering*, 12(2), 169-181.

Vapnik, V. (1998). Statistical learning theory. New York: John Wiley & Sons.

Biography

Luca Abele Piciaccia developed his 25+ years systems engineering career in the Subsea Oil Industry at major oil companies and leading EPC contractors, where he held several positions as Engineering Manager and Commissioning Responsible before becoming Chief Engineer for Subsea Systems Engineering for Major Subsea Projects at Aker Solutions. His education includes Nuclear Engineering at the Politechnic University in Milano, BSc in Engineering Management at CCU and an MBA in Technology Management from Chiefly Business School in Australia. Luca has served 2 terms as president of the Norwegian INCOSE chapter.

Roberto Basili is Associate Professor of Computer Science at the University of Roma, Tor Vergata, where he carries out research and development on Natural Language Processing, Machine Learning, Information Extraction and Retrieval. Currently, he is the coordinator of the SAG@ART group and technical responsible of the AI Laboratory of the Enterprise Engineering Department. His research interests include Representation and Learning of Lexical Knowledge, Information Retrieval and Extraction, as well as Ontology Engineering. He acted as unit technical coordinator within several European Projects (NOMOS, TREVI, ECRAN, NAMIC, PrestoSpace, INSEARCH) on advanced applications of language technologies to human computer interfaces or Web retrieval services. He is author of more than 100 papers on international journals, books and conference proceedings.

Danilo Croce has Ph.D. in Informatics Engineering from the University of Roma "Tor Vergata". Currently, he is a postdoc and he is a member of the SAG@ART group at the Enterprise Engineering Department in the same university. His expertise concerns theoretical and applied Machine Learning in the areas of Natural Language Processing, Information Retrieval and Data Mining. In particular, he is interested in innovative kernels within support vector and other kernel-based machines for advanced syntactic/semantic processing.

Cecilia Haskins entered academia after nearly forty years in industry, including consulting and training in precursor applications of model-based engineering. Her educational background includes a BSc in Chemistry from Chestnut Hill College, and an MBA from Wharton, University of Pennsylvania. She is an Associate Professor at NTNU where she continues her PhD research on innovative applications of systems engineering, with a focus on the oil and gas industry.

Appendix 1 - Physical Quantities

In this section the complete set of 20 Physical Quantities considered in this project are reported.

- **`Acceleration Measure`**
- **`Angle Measure`**
- **`Area Measure`**
- **`Distance Measure`**
- **`Electric Current Measure`**
- **`Energy Measure`**
- **`Flowrate Mass Measure`**
- **`Force Measure`**
- **`Frequency Measure`**
- **`Hardness Measure`**
- **`Power Measure`**
- **`Pressure Measure`**
- **`Rotation Measure`**
- **`Temperature Measure`**
- **`Time Measure`**
- **`Velocity Measure`**
- **`Velocity Angular Measure`**
- **`Volume Measure`**
- **`Weight Measure`**
- **`Percentage Variation`**

Appendix 2 - The MAC classes

In this section the complete set of more that 2,100 MACS classes considered in this project are reported. MACS classes are directly used for ERP processing, from specification to purchase and traceability.

```
1 1000000    EQUIPMENT
      2 1010000    DRILLING EQUIPMENT
            3 1010100    Drilling Machinery, Mud Equipment
                  4 1010101    Mud Handling Equipment
                  4 1010102    Drilling Machinery
            3 1010200    Production Surface Equipment
                  4 1010201    Driller Cabins
                  4 1010202    Driller Control Systems
            3 1010300    Drilling and retrievable Prod. Tools
                  4 1010301    Drilling & Retrievable Production Tools
            3 1010400    Casing, Tubing, Liner, Connectors
                  4 1010401    Ballistic and Perforating Equipment
                  4 1010402    Connectors, Wellhead
            3 1010500    Cementing and Liner Hanger Systems
                  4 1010501    Cement Units
            3 1010600    Fishing and Repair Tools (Drilling)
                  4 1010601    Liner Thrower
            3 1010700    Drilling and Mud Control Instruments
                  4 1010701    Mud Liquid Additive and Control Systems
            3 1010800    Production Well Test and Monitoring
                  4 1010801    Well Monitoring Systems
                  4 1010802    Sample Quills
            3 1010900    Wellhead Equipment, Xmas Trees
            3 1011000    Production String Components
                  4 1011001    Production String Components
            3 1011100    Derricks and Accessories
                  4 1011101    Derricks
            3 1011200    Drill Bits
            3 1011300    BOP and Accessories
                  4 1011301    BOP and Accessories
            3 1011400    Wireline Equipment
                  4 1011401    Wireline Tools
                  4 1011402    Wireline Units
                  4 1011403    Wireline Plugs
                  4 1011404    Wireline Surface Equipment
                  4 1011405    Wireline Cables
                  4 1011406    Wireline Pressure Control Equipment
                  4 1011407    Wireline Pumps and Panels
                  4 1011408    Wireline Containers
                  4 1011409    Wireline Surface Accessories
            3 1011500    Coiled Tubing Tools and Accessories
                  4 1011501    Coiled Tubing Tools
            3 1011600    Down Hole Pressure Control Equipment
                  4 1011601    Downhole Pressure Control Equipment
            3 1011700    Pipe Handling and Lifting Equipment
                  4 1011701    Horizontal Pipe Handling
                  4 1011702    Vertical Pipe Handling
            3 1011800    Subsea Equipment
            3 1019900    Other Drilling and Production Equipment
      2 1020000    HANDLING EQUIPMENT
            3 1020100    Hoist Equipments
                  4 1020101    Davits
                  4 1020102    Monorails
                  4 1020103    Cranes
                  4 1020104    Hoists and Winches
```

4 1020105 Lifting Brakes
3 1020200 Elevators
4 1020201 Personnel Lifts
4 1020202 Goods Lifts
3 1020300 Lifting Equipment
4 1020301 Lifting Jibs
4 1020302 Shackles, Wires and Chains
4 1020303 Lifting Tools and Lift Lugs
4 1020304 Eye Bolts
4 1020305 Blocks, Swivels and Sheaves
4 1020306 Lifting Tables
4 1020307 Crane Adaptors
4 1020308 Lifting Gear Accessories
4 1020309 Sling Sets
3 1020400 Trucks etc.
4 1020401 Forklifts
4 1020402 Automated Guided Vehicles (AGV)
3 1020500 Containers & Baskets
4 1020501 Skids, Transport
4 1020502 Baskets
4 1020503 Containers
3 1020600 Conveyors & Feeders
4 1020601 Screw Conveyors
3 1020700 Packing Equipment and Accessories
4 1020701 Packing Material
4 1020702 Sack Cutting Machines
3 1020900 Jetty Equipment incl. Loading Arms/Hoses
4 1020901 Hose Loading Station
3 1021000 Scaffolding
4 1021001 Scaffolding Tubes
4 1021002 Scaffolding Boards
4 1021003 Scaffolding Couplers
4 1021004 Ladders
4 1021005 Scaffolding Accessories
3 1021100 Tarpaulins
4 1021101 Tarpaulins, Soft Cover
4 1021102 Tarpaulins, Glass Reinforced Polymer
3 1021200 Weighing Equipment and Accessories
4 1021201 Weights
4 1021202 Load Cells
3 1025000 Protective equipment
4 1025001 Caps, Mechanical Protection
4 1025002 Caps, Electrical Protection
4 1025003 Caps, Hydraulical Protection
3 1029900 Other Material and Product Handling Equipment
2 1030000 COMPRESSORS/EXPANDERS/BLOWERS
3 1030100 Centrifugal Compressors
4 1030101 Compressors, Centrifugal/Radial
3 1030200 Reciprocating Compressors
4 1030201 Compressors, Air
4 1030202 Compressors, Nitrogen
4 1030203 Compressors, Hydrogen
4 1030204 Compressors, Diapraghm
4 1030205 Compressors, Single Acting
4 1030206 Compressors, Double Acting
3 1030300 Screw/Rotary Compressors
4 1030301 Compressors, VOC
4 1030302 Compressors, Rotary Screw
4 1030303 Compressors, Rotary Vane
4 1030304 Compressors, Scroll
4 1030305 Compressors, Lobe
4 1030306 Compressors, Liquid Ring
3 1030400 Axial Compressors
4 1030401 Compressors, Axial-Flow
3 1030500 Blowers and Fans
4 1030501 Blowers
4 1030502 Air Fans, Industrial

- 3 1030600 Gas Expanders
  - 4 1030601 Turbo Expanders
- 3 1030700 Jet Compressors
- 3 1030800 Steam Ejectors / Thermocompressors
- 3 1039900 Other Compressors, Blowers and Accessories

2 1040000 PUMPS AND ACCESSORIES

- 3 1040100 Centrifugal and Rotary Pumps
  - 4 1040101 Pumps, Centrifugal
  - 4 1040102 Pumps, Rotary
  - 4 1040103 Pump Shafts
  - 4 1040104 Pump Impellers
  - 4 1040105 Pumps, Seawater Lift
  - 4 1040106 Pumps, Water Injection
- 3 1040200 Reciprocating Pumps
  - 4 1040201 Pumps, Single Acting Reciprocating
  - 4 1040202 Pumps, Double Acting Reciprocating
- 3 1040300 Diaphragm Pumps
  - 4 1040301 Pumps, Membrane
  - 4 1040302 Pumps, Pneumatic Diaphragm
- 3 1040400 Screw Pumps
  - 4 1040401 Pumps, Positive Displacement Screw
- 3 1040500 Submersible Pumps
  - 4 1040501 Pumps, Electric Submersible
- 3 1040600 Multiphase Pumps
  - 4 1040601 Pumps, Helico-Axial Multiphase
  - 4 1040602 Pumps, Twin Screw Multiphase
  - 4 1040603 Pumps, Progressive Cavity Multiphase
- 3 1040700 Liquid Jet Pumps / Eductors
  - 4 1040701 Pumps, Liquid Jet Ventilators
  - 4 1040702 Pumps, Liquid Jet Compressors
  - 4 1040703 Pumps, Liquid Jet
- 3 1040800 Eductors
  - 4 1040801 Pumps, Eductor-Jet
  - 4 1040802 Compressors, Liquid-Jet
- 3 1040900 Slurry Pumps
  - 4 1040901 Pumps, Centrifugal Slurry
  - 4 1040902 Pumps, Lobe Slurry
  - 4 1040903 Pumps, Peristaltic Hose Slurry
- 3 1041000 Hydraulic Pumps
  - 4 1041001 Pumps, Gear
  - 4 1041002 Pumps, Rotary Vane
  - 4 1041003 Pumps, Radial Piston
  - 4 1041004 Pumps, Axial Piston
- 3 1041100 Vacuum Pumps
  - 4 1041101 Pumps, Liquid Ring Vacuum
- 3 1041200 Pump Skids
  - 4 1041201 Skids, Pumps
- 3 1049900 Other Pumps and accessories

2 1050000 DRIVERS AND ACCESSORIES

- 3 1050100 Electric Motors / Drivers
  - 4 1050101 Electrical Motors
  - 4 1050102 Electrical Motor Parts
- 3 1050200 Gas Turbines
  - 4 1050201 Auxiliary Power Units (APUs)
  - 4 1050202 Gas Turbines for Power Generation
  - 4 1050203 Gas Turbines for Mechanical Drive
- 3 1050300 Steam Turbines
- 3 1050400 Oil/Diesel Engines
  - 4 1050401 Engines, Diesel
- 3 1050500 Air/Gas Engines
- 3 1050600 Hydraulic Motors
  - 4 1050601 Motors, Gear
  - 4 1050602 Motors, Rotary Vane
  - 4 1050603 Motors, Radial Piston
  - 4 1050604 Motors, Axial Piston
- 3 1050700 Gear Boxes, Gear Units, Couplings
  - 4 1050701 Gear boxes

4 1050702 Clutches
4 1050703 Couplings
4 1050704 Pinion Gears
3 1050800 Submersible Motors
3 1050900 Propulsion Units and Accessories
3 1059900 Other Drivers and accessories
2 1060000 HEATERS/FURNACES/ BOILERS ETC
3 1060100 Direct fired Heaters/Boilers
4 1060101 Heaters/Boilers, Gas Burner
4 1060102 Heaters/Boilers, Oil-Fired
3 1060200 Steam Boilers
3 1060300 Electric Heaters
4 1060301 Heaters, Electric Radiative
4 1060302 Heaters, Electric Convection
4 1060303 Heaters, Electric Fan
4 1060304 Heaters, Electric Oil-Fired
3 1060400 Stacks, Flares and Accessories
4 1060401 Flares and Accessories
4 1060402 Flare Tips
3 1069900 Other Heaters, Furnaces, Boilers and Accessories
2 1070000 HEAT EXCHANGERS/HEAT TRANSFER EQUIPMENT
3 1070100 Shell and Tube Heat Exchangers
4 1070101 Heat Exchangers, U-tube
4 1070102 Heat Exchangers, Straight-Tube
3 1070200 Plate Heat Exchangers (PHE)
4 1070201 PHE Intermittent Corrugation
4 1070202 PHE Chevron Corrugation
3 1070300 Air Coolers
4 1070301 Air Coolers, Brine
4 1070302 Air Coolers, Evaporative
3 1070400 Waste Heat Recovery Units
3 1070500 Kettle Reboiler
3 1070600 Spiral Heat Exchangers
4 1070601 Flow Exchangers, Counter-Current
4 1070602 Flow Exchangers, Spiral Flow/Cross
4 1070603 Flow Exchangers, Distr. Vapour/Spiral
4 1070604 Heat Exchangers, Hydr. Liquid to Liquid
4 1070605 Heat Exchangers, Hydr. Liquid to Air
4 1070606 Heat Exchangers, Hydr. Liquid to Water
3 1070700 Transfer Line Exchangers
4 1070701 Transfer Line Exchangers, Multiple Tube
4 1070702 Transfer Line Exchangers, Linear
3 1070800 Printed Circuit Heat Exchangers
4 1070801 Heat Exchangers, Printed Circuit
3 1079900 Other Heat Exchangers/Heat Transfer Equipment
2 1080000 TANKS/VESSELS/COLUMNS ETC.
3 1080100 Storage Tanks incl. Spheres
4 1080101 Tanks, Atmospheric
4 1080102 Tanks, Gas Cylinder
4 1080103 Tanks, Bulk
4 1080104 Tanks, Septic
4 1080105 Tanks, External Flooting Roaf
4 1080106 Reservoir, Hydraulic
4 1080107 Reservoir, Pneumatic
3 1080200 Pressure Vessels, Drums, Accumulators
4 1080201 Pressure Vessels
4 1080202 Pulsation Dampeners
4 1080203 Expansion Bellows
4 1080204 Autoclaves
4 1080205 Separation Vessels
4 1080206 Accumulators
4 1080207 Pressure Drums
3 1080300 Columns and Accessories
4 1080301 Silos
3 1080400 Tank/Vessel/Column Internals
4 1080401 Distillation Colums
4 1080402 Vessel Internals

3 1089900 Other Tanks, Vessels, Columns and Accessories
2 1090000 ELECTRICAL EQUIPMENT AND MATERIALS
3 1090100 Generators, Power Sources, Units
4 1090101 Electrical Power Units (EPU)
4 1090102 Main Power Generators
4 1090103 Emergency & Essential Generators
3 1090200 Transformers and Accessories
4 1090201 Transformers - High Voltage (HV)
4 1090202 Transformers - Low Voltage (LV)
3 1090300 Switch/Control Equipm, Plugs, Connectors
4 1090301 Control Equipment
4 1090302 Printed Circuit Boards (PCB/PWA)
4 1090303 Electrical Plugs and Connectors
4 1090304 Slip rings
4 1090305 Switchgears
3 1090400 Rectifiers, Inverters and Converters
4 1090401 Rectifiers
4 1090402 Inverters
4 1090403 Converters
3 1090500 Distribution Equipment and Materials
4 1090501 Topside Power Distribution Unit
4 1090502 Junction Boxes
4 1090503 Subsea Power Distribution Unit
4 1090504 Variable Speed Drive (VSD)
4 1090505 Circuit Breakers
3 1090600 Lamps, Lighting Fixtures and Equipment
4 1090601 Lamps
4 1090602 Lamp Fixtures
3 1090700 Cables, Cords, Wires and Accessories
4 1090701 Cables, Power Supply
4 1090702 Electrical Cable Accessories
4 1090703 Jumpers, Electrical / Optical Subsea
3 1090800 Cathodic Protection Equipment
4 1090801 Cathodes
3 1090900 Transit and Glands
4 1090901 Cable Transits
3 1091000 Electrical Connectors - Subsea
4 1091001 Feedthrough Systems, Electric/Optic
4 1091002 Connectors, Electrical Subsea
3 1091100 Cable Racks and Trays
4 1091101 Cable Enclosures
4 1091102 Cable Trays
4 1091103 Cable Spools
3 1099900 Other Electrical Equipment and Materials
2 1100000 INSTRUMENTATION AND PROCESS CONTROL
3 1100100 Pressure Instruments
4 1100101 Pressure Transmitters, Gauge
4 1100102 Pressure Transmitters, Differential
4 1100103 Pressure Transmitters, Absolute
4 1100104 Pressure Gauges
4 1100105 Sensors, Pressure & Temperature (PTT)
3 1100200 Temperature Instruments
4 1100201 Temperature Transmitters
4 1100203 Thermostats
4 1100202 Temperature Gauges
4 1100204 Temperature Instruments, Accessories
3 1100300 Level Instruments
4 1100301 Level Transmitters, Differential
4 1100302 Level Transmitters, Micro Wave
4 1100303 Level Transmitters, Radar
4 1100304 Level Transmitters, Displacement
4 1100305 Nucleonic Level Measurement Systems
4 1100306 Level Gauges
3 1100400 Flow Instruments / Flow Meters (FM)
4 1100401 Flow Meters, Mechanical
4 1100402 Flow Meters, Pressure-Based
4 1100403 Flow Meters, Optical

4 1100404 Flow Meters, Open Channel
4 1100405 Flow Meters, Thermal Mass
4 1100406 Flow Meters, Vortex
4 1100407 FM, Electrom., Ultrasonic, Coriolis
4 1100408 Flow Meters, Laser Doppler
4 1100409 Flow Meters, Multi-Phase (MPFM)
4 1100410 DHTP Cards
3 1100500 Fire/Smoke/Gas/Heat Detection Instrum.
4 1100501 Detectors, Fire
4 1100502 Detectors, Smoke
4 1100503 Detectors, Gas
4 1100504 Detectors, Heat
4 1100505 Manual Call Point (MAC)
3 1100600 Instrument Tubing and Fittings
4 1100601 Instrument Tubing
4 1100602 Instrument Fittings
3 1100700 Junction Boxes
3 1100800 Process Control and Monitoring Systems
4 1100801 Automation and Control Systems
4 1100802 Condition Monitoring Systems
3 1100900 Metering Equipment and Systems
4 1100901 Logging Equipment
4 1100902 Sensors, Vibration
4 1100903 Fiscal Metering
3 1101000 Detectors and Analysers
4 1101001 Detectors, Water Ingress
4 1101002 Detectors, Leak
4 1101003 Detectors, Sand
4 1101004 Analysers, PMI
4 1101005 Analysers, Gas
4 1101006 Analysers, Hydrocarbon in Water
4 1101007 Analysers, Water in Oil
4 1101008 Analysers, Redox
4 1101009 Analyers, Oxygen in Vapour
4 1101010 Analysers, Hydrocarbons
3 1101100 Control Panels and Control Stations
4 1101101 Control Panels
4 1101102 Control Stations
4 1101103 Wireless Communication Units
3 1101200 Sonar/Radar/DP/Positioning Systems
4 1101201 DP-Positioning Systems
4 1101202 Motion Reference Units (MRU)
4 1101203 Approach and Mooring Solutions
3 1101300 Corrosion/Erosion Monitoring Systems
4 1101301 Detectors, Subsea Corrosion
4 1101302 Detectors, Subsea Erosion
3 1101400 Simulator Systems
4 1101401 Simulators, Sensor
4 1101402 Simulators, Process
3 1101500 Acoustic Equipment
4 1101501 Acoustic Imaging Equipment
4 1101502 Sonaring Equipment
3 1101600 Acoustic Telemetry Systems
4 1101601 Acoustic Telemetry Systems
3 1101700 Meteorological Instruments/Equipment
4 1101701 Meteorological Instruments
3 1105000 Electronic Accessories
4 1105001 Electronic Components
4 1105002 CPDU Internals
3 1109700 Different Alarm/Control Systems
4 1109701 Alarm Systems
3 1109800 Instrument and Communication Cables
4 1109801 Cables, Control
4 1109802 Cables, Communication
4 1109803 Cables, Fibre Optical
4 1109804 Cables, Coaxial
3 1109900 Other Instrument / Communication and Process Control

Equipment / Materials

- 2 1110000 MISCELLANEOUS MECHANICAL EQUIPMENT
  - 3 1110100 Process treatment Equipment
    - 4 1110101 Hydrocyclons
    - 4 1110102 Sandcyclons
    - 4 1110103 Centrifuges
    - 4 1110104 Scrubbers
    - 4 1110105 Separator PV's and Internals
  - 3 1110200 Regen./Non-Regen. Filters, Strainers
    - 4 1110201 Strainers
    - 4 1110202 Filters, Regenerative
    - 4 1110203 Filter Cleaning (acid wash)
    - 4 1110204 Membranes
    - 4 1110205 Filter Press
    - 4 1110206 Filters, Non-Regenerative
  - 3 1110300 Mixers, Agitators, Blenders, Feeders
    - 4 1110301 Agitators (motor driven)
    - 4 1110302 Feeders, Cell
    - 4 1110303 Lube Oil Accesoiries
  - 3 1110400 Hydraulic Units and Accessories
    - 4 1110401 Hydraulic Power Unit (HPU)
    - 4 1110402 Hydraulic Cylinders
    - 4 1110403 Hydraulic Accessories
    - 4 1110404 Hydraulic Valves
    - 4 1110405 Hydraulic Hoses
  - 3 1110500 Fresh Water Treatm./Desalination Systems
    - 4 1110501 Fresh Water Makers
    - 4 1110502 Deaeration Systems
  - 3 1110600 Gas Treatment Units and Accessories
    - 4 1110601 Nitrogen Gas Generators
  - 3 1110700 Press. Red. Stations for Gas Distr.
  - 3 1119900 Other Mechanical Equipment and Accessories
- 2 1120000 AIR CONDITIONING EQUIPMENT
  - 3 1120100 HVAC System Packages
    - 4 1120101 HVAC system / packages
  - 3 1120200 Cooling and Refrigeration Units
    - 4 1120201 Cooling Systems
  - 3 1120300 Humidifiers, Driers (Dryers) etc.
    - 4 1120301 Humidifiers
    - 4 1120302 Driers
  - 3 1120400 Air Fans
    - 4 1120401 Air Fans, Domestic
  - 3 1120500 Air Filters, Coalescers and Accessories
    - 4 1120501 Demisters
    - 4 1120502 Air Filters
    - 4 1120503 Air Filter Accessories
  - 3 1120600 Dampers and Accessories
    - 4 1120601 Dampers
  - 3 1120700 Ducting etc.
  - 3 1120800 Heating Coils/Units
  - 3 1129900 Other HVAC Equipment and Accessories
- 2 1130000 MARINE / DIVING AND PIPELINE EQUIPMENT
  - 3 1130100 Anchoring/Buoying Equipment
    - 4 1130101 Anchor Chains
    - 4 1130102 Buoyancy Elements
    - 4 1130103 Anchors
  - 3 1130200 ROV's, ROV Tools and Accessories
    - 4 1130201 Launch and Recovery Systems (LARS)
    - 4 1130202 Submarines
    - 4 1130203 ROV's and Accessories
  - 3 1130400 Ships Gears
  - 3 1130500 Turrets
  - 3 1130600 Personal Diving Equipment
  - 3 1130700 Vessel based Diving Equipment
  - 3 1132000 Pipeline Equipment and Connectors
    - 4 1132001 Connectors, Tie-Back
  - 3 1132100 Pig Launchers/Transm./Traps/Receivers

4 1132101 Pig Launchers
4 1132102 Transmitters
4 1132103 Traps
4 1132104 Receivers
3 1132200 Pigging Equipment and Accessories
4 1132201 Pigging Systems
3 1132300 Through Flowline Systems and Tools
3 1139900 Other Marine / Diving and Pipeline Equipment and Accessories
2 1140000 COMPUTER / COMMUNICATION HARD/SOFT-WARE
3 1140100 Computer Hardware
4 1140101 Notebooks
4 1140102 Desktop Computers
4 1140103 Mobile Workstations
4 1140104 Desktop Workstations
4 1140105 Monitors
3 1140200 IT Infrastructure
4 1140201 Servers
4 1140202 Data Rooms
4 1140203 Storage & Backup
4 1140204 Network Equipment
4 1140205 IT Infrastructure Accessories
3 1140300 Computer Hardware - Mainframes
3 1140400 Computer Peripherals
4 1140401 Printers & Scanners
4 1140402 Multi-Functional Devices (MFD)
4 1140403 Computer & Printer Accessories
3 1140500 Project and Planning (IT) Applications
4 1140501 Applications - Planning
4 1140502 Applications - Risk Management
4 1140503 Applications - Quality/Performance
4 1140504 Applications - Document Management
4 1140505 Applications - Project Systems
4 1140506 Applications - Resource Management
4 1140507 Project & Planning Applications - Cost Management
3 1140600 IT Application Development Tools
4 1140601 Tools - MS Development
4 1140602 Tools - Java
4 1140603 Tools - Testing
4 1140604 Tools - WEB development
4 1140605 Tools - Data Modelling
4 1140606 Tools - Programming Tools
4 1140607 Tools - Desktop d/b
3 1140700 Business Management (IT) Applications
4 1140701 Product Lifecycle Management (PLM)
4 1140702 Business Intelligence (BI)
4 1140703 Finance
4 1140704 Supply Chain Management
4 1140705 Manufacturing
4 1140706 Logistics
4 1140707 Aftermarket & Customer Service
4 1140708 Governance, Risk and Compliance
4 1140709 Sales
4 1140710 Human Resources (HR)
4 1140711 Plant Maintenance
4 1140712 Service Management
4 1140713 Survey and Data Management
3 1140800 Engineering (IT) Applications
4 1140801 General Engineering
4 1140802 Engineering Process
4 1140803 Structure & Material Tech.
4 1140804 Safety & Environment
4 1140805 Electrical & Instrumentation
4 1140806 Piping & Layout
4 1140807 Engineering Register
4 1140808 Change Control Systems
3 1140900 Design Automation (IT) Applications

4 1140901 Plant Design
4 1140902 Product Design
4 1140903 Applications - Automation
4 1140904 Product Data Management (PDM)
4 1140905 Visualization
4 1140906 Surveillance
3 1141000 General Office (IT) Applications
4 1141001 Standard Office Products
4 1141002 Communication & Collaboration
4 1141003 Office Accessories
4 1141004 Learning
4 1141005 Inhouse IT
4 1141006 Language
3 1141100 IT Systems
4 1141101 IT Systems - Security Software
4 1141102 IT Systems - Operating Systems
4 1141103 IT Systems - Database
4 1141104 IT Systems - Printing
4 1141105 IT Systems - Middelware
4 1141106 IT Systems - System Administration
4 1141107 IT Systems - System Utilities
3 1142000 Mobility & (Tele)Communication Equipment
4 1142001 Telecommunication Equipment
4 1142002 Radio systems
4 1142003 Telecommunication Monitoring Systems
4 1142004 PABX Telephone Systems
4 1142005 Telex / Telefax
4 1142006 Mobile Phones
4 1142007 IP fixed phones
4 1142008 Tablets
4 1142009 Communication Accessories
3 1142100 Audio/Video Equipment & Accessories
4 1142101 Cameras
4 1142102 Audio Equipment
4 1142103 Video Equipment
4 1142104 Monitors Public Display
4 1142105 Administration Panels
4 1142106 Codecs
4 1142107 Standard VC Solutions
4 1142108 Projectors
4 1142109 VC Accessories
3 1142200 Personal Communication Equipment
3 1149900 Other Computer and Communication Equipment and Accessories
2 1150000 LABORATORY / MEDICAL EQUIPMENT
3 1150100 General Lab. Apparatuses and Requisites
4 1150101 Laboratory Equipment
4 1150102 Metric Tapers
3 1150300 Medical Equipment and Supplies
4 1150301 Pharmaceutical Products
4 1150302 Medical Equipment
3 1159900 Other Laboratory and Medical Equipment and Accessories
2 1160000 SAFETY/PROTECTION/SECURITY/FIREFIGHTING
3 1160100 Security Equipment and Accessories
4 1160101 Access & Control Systems
3 1160200 Safety/Prot. Equipm., Life Boats/Rafts
4 1160201 Safety Equipment
4 1160202 Protection Equipment
4 1160203 MOB- / Life Boats
3 1160300 Fire Fighting Equipment/Products
4 1160301 Fire Water Hydrants
4 1160302 Hose Reel Cabinet
4 1160303 Fire Extinguishers
4 1160304 Sprinklers and Accessories
4 1160305 Fire and Explosion Proof Doors
4 1160306 Pumps, Firewater
3 1160400 Escape Tools/Equipment

4 1160401 Helicopter Rescue Equipment
4 1160402 Escape Chute Equipment
3 1160500 Fire/Gas Detection/Protection Systems
4 1160501 Fire/Gas Detection Systems
3 1160600 Inert Gas Systems/Nitrogen
4 1160601 Inert Gas Systems/Nitrogen
3 1165000 Public Address System
4 1165001 PA Speakers
4 1165002 Flash Beacon Lights
4 1165003 Public Address and Alarm Systems
3 1169900 Other Safety / Security / Firefighting Equipment and Accessories
2 1170000 ENVIRONMENTAL EQUIPMENT/PRODUCTS
3 1170100 Waste Water Disposal/Recovery Equipment
4 1170101 Sewage Treatment Systems
4 1170102 Produced Water Treatment Equipment
3 1170200 Oil Recovery Equipment and Accessories
4 1170201 Sulphate Removal Equipment
3 1170300 Waste Gas Treatment/Recovery Equipment
4 1170301 Vapour Recovery Equipment
4 1170302 Flare Gas Recovery Equipment
3 1170400 Noise Abatement Equipment
3 1170500 Asbestos Removal Equipment
3 1179900 Other Environmental Equipment / Products
2 1180000 PACKAGES - CONSTRUCTION/OUTFITTING
3 1180300 Chemical Injection Systems
4 1180301 Chemical Injection Systems
4 1180302 Skids, Chemical Injection Systems
3 1180400 Subsea Packages
3 1180500 Helideck Systems/Packages
3 1180600 Steel/Metal/Composite Outfitting
4 1180601 Railing
3 1180700 Prefabricated Spools, Pipework etc.
3 1189900 Other Industrial Equipment, Packages and Units
2 1190000 SERVICE STATION EQUIPMENTS / MATERIALS
3 1190100 Shop Fittings
3 1190200 Car Wash Equipment and Accessories
3 1190300 Canopies and Accessories
3 1190400 Service Station Signs and Accessories
3 1190500 Service Station Pumps and Accessories
3 1190600 Service Station Payment Terminals
3 1190700 Service Station Tanks and Accessories
3 1190800 Food Products for Resale
3 1190900 Fast Food Products for Resale
3 1191000 Non-food Products for Resale
3 1191100 Home Entertainment Products for Resale
3 1191200 Newspapers and Magazines (for Resale)
3 1199900 Other Service Station Equipment and Materials
2 1500000 ALUMINIUM PRODUCTION EQUIPMENT
3 1500100 Anode Baking Furnaces and Accessories
3 1500200 Aluminium Electrolysis Cells
3 1500300 Casting Machines and Accessories
3 1500400 Ceramic Filters
3 1500500 Refractory Material and Cathode Carbon
4 1500501 Activated Carbon
3 1500600 Steel for Electrolytic Reduction Cells
3 1500700 Extrusion Billets
3 1500800 Aluminium Coils
3 1500900 Silica Sand
3 1501000 Foundry Alloys
4 1501001 Aluminium Foundry Alloys
3 1501100 Fasteners
3 1501200 Gaskets for Building Systems
3 1501300 Fittings for Building Systems
3 1501400 Accessories for Building Systems
3 1501500 Paints and Lacquers
3 1501600 Papers and Films

1 2000000 MATERIALS & COMPONENTS
2 2010000 PIPES / TUBES / HOSES AND FITTINGS
3 2010100 Seamless Pipes and Tubes
4 2010101 Tubing, Seamless
4 2010102 Pipes, Seamless
3 2010200 Welded Pipes and Tubes
4 2010201 Flow Loops, Welded
4 2010202 Tabulars, Welded
4 2010203 Spools, Welded
4 2010204 Pipes, Welded
4 2010205 Tubes, Welded
3 2010300 Non-metal Pipes and Tubes
4 2010301 Pipes, GRE (Glassfiber Reinforced Epoxy)
4 2010302 Tubes, GRE (Glassfiber Reinforced Epoxy)
4 2010303 Rods, Carbon Fibre
3 2010400 Hoses
4 2010401 Hoses, Steel
4 2010402 Hoses, Rubber
3 2010500 Flanges, Elbows, Fittings and Supports
4 2010501 Flanges
4 2010502 Pipe Fittings
4 2010503 Clamps and Pipe Supports
4 2010504 Tube Flanges
4 2010505 Flow Loops Assemblies
4 2010506 Adapters
4 2010507 Couplers
4 2010508 Self Sealing Clamps
4 2010509 Elbows and Goose Necks
4 2010510 Hangers
4 2010511 Carriers
4 2010512 Hose Reels
3 2010600 Gaskets, Jointings, Packings
4 2010601 Gaskets
4 2010602 Jointings
4 2010603 Packings
4 2010604 Tightenings
3 2010700 Non-Metal Pipe Equipments
4 2010701 GRE, GRP, Thermo -hardened -plastic -set
3 2010800 Insulated Pipes, Tubes, Hoses
4 2010801 Insulated Pipes, Tubes, Hoses
3 2010900 Flexible Pipes
4 2010901 Pipes, Flexible
3 2019900 Other Pipes, Tubes, Hoses, Hubs, Pipe Penetrators and Fittings etc.
2 2020000 VALVES AND ACCESSORIES
3 2020200 Check Valves
4 2020201 Check Valves, Shuttle
4 2020202 Check Valves, Swing
4 2020203 Check Valves, Wafer
4 2020204 Check Valves, Piston
4 2020205 Check Valves, Axial Flow
4 2020206 Check Valves, Non-Slam
3 2020300 Directional Valves
4 2020301 Directional Valves
3 2020400 Pressure Control Valves
4 2020401 Pressure Control Valves, Relief
4 2020403 Pressure Control Valves, Reducing
4 2020404 Pressure Control Valves, Sequence
4 2020405 Pressure Control Valves, Counterbalance
4 2020406 Pressure Control Valves, Unloading
3 2020500 Subsea Valves
3 2020600 Solenoid Valves
4 2020601 Solenoid Valves
3 2020700 Actuators and Valve Instrumentation
4 2020701 Actuators, Control
4 2020702 Actuators, On-Off
4 2020703 Actuators, Assembly

4 2020704 Actuators, Anti-Surge
4 2020705 Actuators, Rotating
4 2020706 Actuators, Linear
4 2020707 Actuators, Twist
3 2020800 Ball Valves
4 2020801 Ball Valves, Top entry
4 2020802 Ball Valves, End entry
4 2020803 Ball Valves, Split body
4 2020804 Ball Valves, Side Entry
4 2020805 Ball Valves, Kelly
3 2020900 Butterfly Valves
4 2020901 Butterfly Valves
4 2020902 Chemical Injection Valve, Retrievable
3 2021000 Diaphragm Valves
4 2021001 Diaphragm Valves
3 2021100 Gate Valves
4 2021101 Gate Valves, Through Conduit
4 2021102 Gate Valves, Solid
4 2021103 Gate Valves, Split
4 2021104 Gates and Seats
4 2021105 Gate Valves, Slab
4 2021106 Gate Valves, Double Expanding
3 2021200 Globe Valves
4 2021201 Globe Valves
3 2021300 Needle Valves
4 2021301 Needle Valves, Manual
3 2021400 Plug Valves
4 2021401 Plug Valves
3 2021500 Rotary disk valves
4 2021501 Rotary Disk Valves, Manual
4 2021502 Rotary Disk Valves, Hydraulic
3 2021600 Choke Valves
4 2021601 Choke Valves, Insert Retrievable
4 2021602 Chokes, Fixed Type
4 2021603 Chokes, Tooling
3 2021700 Bursting Disc Valves
4 2021701 Rupture Disc Valves
3 2029900 Other Valves and Accessories
2 2030000 STEEL / METAL MATERIALS
3 2030100 Steel / Metal Plates
4 2030101 Plates, Raw Metal
4 2030102 Plates, Perforated Metal
4 2030103 Sheets, Metal
4 2030104 Foil, Metal
3 2030200 Steel / Metal Bars
4 2030201 Bars, Round Metal
4 2030202 Bars, Square Metal
4 2030203 Bars, Rectangular Metal
3 2030300 Steel / Metal Profiles and Sections
4 2030301 Weldments, Metal
4 2030302 Frames, Metal
4 2030303 Brackets, Metal
4 2030304 Channels, Metal
4 2030305 Universal Beams, Metal
4 2030306 Angles, Metal
4 2030307 Joists, Metal
4 2030308 Columns, Metal
4 2030309 Bearing Piles, Metal
4 2030310 Tee Bars, Metal
4 2030311 Hollow Sections, Metal
3 2030400 Steel / Metal Forgings and Castings
4 2030401 Forgings, Block Metal
4 2030402 Forgings, Ring Metal
4 2030403 Forgings, Ring Rolled Metal
4 2030404 Forgings, Open Die Metal
4 2030405 Forgings, Closed Die Metal
4 2030406 Castings, Metal

```
            4 2030407    Hot isostatic pressing (HIP)
      3 2030500    Steel / Metal Gratings
            4 2030501    Gratings, Press Locked Metal
            4 2030502    Gratings, Heavy Duty Metal
            4 2030503    Gratings, Press Welded Metal
            4 2030504    Gratings, Offshore Metal (O-SP-SS)
            4 2030505    Gratings, Mesh
            4 2030506    Gratings, Fiber-Glass
      3 2030600    Metal Wires and Chains incl. Accessories
            4 2030601    Chain, Side Welded Metal
            4 2030602    Chain, Twisted Link Metal
            4 2030603    Chain, Spiked Metal
            4 2030604    Chain, Single Jack Metal
            4 2030605    Chain, Proof Coil Side Welded Metal
            4 2030606    Chain, Metal Accessoiries
            4 2030607    Wire Rope, Stainless Steel
            4 2030608    Wire Rope, Accessoiries
            4 2030609    Springs and Spirals, Metal
      3 2030700    Steel / Metal Fasteners
            4 2030701    Metal Bolts
            4 2030702    Metal Screws
            4 2030703    Metal Clips / Clasps / Staples / Rings
            4 2030704    Metal Inserts
            4 2030705    Metal Keys
            4 2030706    Metal Nuts
            4 2030707    Metal Pins
            4 2030708    Metal Rivets
            4 2030709    Metal Rods
            4 2030710    Metal Washers
            4 2030711    Metal Stud
      3 2030800    Steel / Metal Bearings and Seals
            4 2030801    Seals, Standard Metal
            4 2030802    Seals, Custom Metal
            4 2030803    Bearings, Metal
            4 2030804    Seal Kits, Metal
            4 2030805    Wipers, Metal
            4 2030806    Seal Retainers, Metal
      3 2030900    Anodes
            4 2030901    Anodes, Aluminium
            4 2030902    Anodes, Zinc
            4 2030903    Anodes, Magnesium
            4 2030904    Anodes, Extruded Magnesium
      3 2031000    Welding Consumables
            4 2031001    Wires, Hardfacing Cored
            4 2031002    Wires, Cladding Cored
            4 2031003    Wires, Aluminium
            4 2031004    Wires, Thermal Arc Spraying Cored
            4 2031005    Wires, Joining Cored
            4 2031006    Welding Electrodes
            4 2031007    Flux Cored TIG
      3 2032000    Doors, Manholes, Hatches
            4 2032001    Doors
            4 2032002    Manholes
            4 2032003    Hatches
      3 2039900    Other Steel and Metal Materials
2 2040000    NON-METAL MATERIALS
      3 2040100    Non-Metal Plates
            4 2040101    Plates, Raw Non-Metal
            4 2040102    Plates, Perforated Non-Metal
            4 2040103    Sheets, Non-Metal
            4 2040104    Foil, Non-Metal
      3 2040200    Non-Metal Bars
            4 2040101    Bars, Raw Non-Metal
      3 2040300    Non-Metal Profiles and Sections
            4 2040301    Profiles, Plastic
            4 2040302    Profiles, PVC
            4 2040303    Profiles, Composite
```

4 2040304 Profiles, Rubber
3 2040400 Non-Metal Forgings and Castings
4 2040401 Castings, Non-Metal
3 2040500 Non-Metal Gratings
4 2040501 Gratings, Non-metal
3 2040600 Non-Metal Ropes & Chains
4 2040601 Ropes, Non-Metal
4 2040602 Ropes, Mooring Polyester
4 2040603 Lashing Equipment, Non-Metal
3 2040700 Non Metal Fasteners
4 2040701 Fasteners, Non Metal Bolts
4 2040702 Fasteners, Non Metal Nuts
4 2040703 Fasteners, Non Metal Washers
3 2040800 Non Metal Bearings and Seals
4 2040801 Seals, Standard Non-Metal
4 2040803 Seals, Custom Non-Metal
4 2040802 Bearings, Non-Metal
4 2040804 Wipers, Non-Metal
4 2040805 Seal Kits, Non-Metal
3 2049900 Other non-Metal Materials
2 2050000 AUTOMOBILES AND ASSOCIATED EQUIPMENT
3 2050100 Trucks and Associated Equipment
3 2050200 Trailers and Associated Equipment
3 2050300 Vans, Lorries and Associated Equipment
3 2050400 Passenger Cars and Associated Equipment
3 2050500 Special Vehicles and Associated Equipm.
3 2050600 Tyres/Batteries/Car Accessories
3 2059900 Other Automobiles and Associated Equipment
2 2060000 DOWNSTREAM (PACKING MATERIALS)
3 2060100 Drums
4 2060101 Drums - Packing
3 2060200 Bottles and Cans
3 2060300 Fluid Bags, Small Containers
4 2060301 Fluidizing Canvas
3 2060400 Boxes (various Materials)
3 2060500 Pallets
4 2060501 Pallets, Wooden
4 2060502 Pallets, Plastic
3 2069900 Other Packing Materials
2 2070000 CHEMICALS / OILS / PAINTS
3 2070100 Primers, Coatings and Paints
4 2070101 Primer
4 2070102 Coating
4 2070103 Paint
3 2070200 Petroleum Products
4 2070201 Hydraulic Fluids
4 2070202 Test Fluids
4 2070203 Oils
4 2070204 Fuels
3 2070300 Mud and Associated Additives
3 2070400 Abrasives/Polishes/Compounds/Adhesives
4 2070401 Adhesives and Epoxies
4 2070402 Polishes
4 2070403 Compounds
4 2070404 Abrasives
3 2070500 Industrial Gases
4 2070501 Gas, Nitrogen
4 2070502 Gas, Oxygen
4 2070503 Gas, Argon
4 2070504 Gas, Helium
4 2070505 Gas, Propene
3 2070600 Cleaning Products
3 2070700 Additives, Inhibitors etc.
3 2070800 Lubricants
4 2070801 Lubrication Oil
4 2070802 Lubrication Grease
3 2070900 Production/Process/Drilling Chemicals

```
            4 2070901    Explosives
            4 2070902    Production Control Fluids
            4 2070903    Metal Powders
      3 2071000    Laboratory Chemicals
      3 2071100    Catalysts
            4 2071101    Biochemicals
      3 2071200    Completion Fluids and Additives
      3 2071300    Cement and Associated Additives
      3 2079900    Other Chemicals, Oils and Paints
2 2080000    INSULATION / REFRACTORY MATERIALS
      3 2080100    Fire Protection Materials
            4 2080101    Fire Protection Materials for Structures
            4 2080102    Fire Protection Materials for Piping
      3 2080200    Thermal Insulation Materials
            4 2080201    Conservation Materials, Heat
            4 2080202    Conservation Materials, Cold
            4 2080203    Condensation Prevention Materials
            4 2080204    Frost Protection Materials
            4 2080205    Personal Protection Materials
      3 2080300    Acoustic "Insulation      Materials
            4 2080301    Acoustic Insulation Materials
      3 2089900    Other Insulation and Refractory Materials
2 2090000    CIVIL BULK MATERIALS
      3 2090100    Cement
      3 2090200    Sand and Aggregates
      3 2090300    Reinforcing Materials
            4 2090301    Reinforcement Pipes
      3 2090400    Inserts incl. Anchors, Bolts etc.
      3 2090500    Protection Materials
      3 2099900    Other Civil Bulk Materials
2 2100000    ARCHITECTURAL / BUILDING MATERIALS
      3 2100100    Sanitary Equipment
            4 2100101    Drains
      3 2100200    Laundry and Galley/Kitchen Equipment
            4 2100201    Galleys
      3 2100300    Floorings and Wall / Ceiling Panels
            4 2100301    Cabins and Modules
            4 2100302    Walls
            4 2100303    Emergency Corridors
            4 2100304    Floors
      3 2100400    Doors, Windows and Shutters
            4 2100401    Doors
            4 2100402    Windows
            4 2100403    Shutters
      3 2100500    Textiles and Wall Papers
      3 2100600    Warehouse Equipment and Supplies
            4 2100601    Warehouse Equipment
            4 2100602    Warehouse Supplies
      3 2100700    Catering Equipment
      3 2100800    Signs and Boards
            4 2100801    Marking (Subsea)
            4 2100802    Warning and Safety Signs
            4 2100803    Office Signs
            4 2100804    Marking (Top-side)
      3 2109900    Other Building Materials
2 2110000    WORKSHOP- AND HANDTOOLS
      3 2110100    Workshop Equipm., Tools and Accessories
            4 2110101    In-house Developed Tools
            4 2110102    Technical Tapes
            4 2110103    Working Clothes
            4 2110104    Industrial Machines
            4 2110105    Cutting Tools
            4 2110106    Ceramic Magnets
      3 2110200    Welding Equipment and Accessories
            4 2110201    Welding Equipment
      3 2110300    Cleaning Equipment and Accessories
            4 2110301    HP Wash Package
```

3 2110400 Mechanical Handtools
4 2110401 General Manufacturing Tools
4 2110402 Pipe Handling Tools
3 2110500 Electrical Handtools
4 2110501 Electrical Workshop Tools
3 2110600 Test and Measuring Instruments/Equipment
4 2110601 Test Stand
4 2110602 Test & Flushing Equipment
4 2110603 Measuring Equipment
4 2110604 Probes
2 2120000 OFFICE ADMIN. PRODUCTS AND CONSUMABLES
3 2120100 Office Supplies/Stationery
4 2120101 Office Supplies / Stationery
3 2120200 IT Consumables
3 2120300 Pre-printed Stationery
3 2120400 Food Products and Beverages
3 2120500 Kitchenware and Tableware
3 2120600 Cleaning/Sanitary Supplies
3 2120700 Furniture
4 2120701 Cabin Chairs
4 2120702 Office Furniture
3 2120800 Office Equipment
4 2120801 Conference Room Equipment
3 2120900 Sports and Leisure Equipment
3 2121000 Newspapers, Magazines, Books - Own use
3 2129900 Other Office Administration Products and Consumables
2 2500000 RAW MATERIALS FOR ANODES AND ALUMINIUM
3 2500100 Coke, Pitch, Prebaked Anodes and Paste
3 2500200 Alumina, Fluoride and Soda
3 2500300 Alloying Elements/Additives
2 2990000 DIFFERENT PRODUCTS / EQUIPMENT AND MATERIALS
3 2990200 Packing, Crating and Boxing Materials
4 2990201 Bags, Bottle, Cans, Caps, IATA, IMDG
3 2990300 Consumables for Testing / Inspection
4 2990301 X-ray Film
3 2990400 Uniforms and various types of Clothing
4 2990401 Footwear, Gloves, Helmets, Uniform
3 2990500 Scale Models (Ships, Rigs, Plants,etc.)
4 2990501 Plants, Rigs, Ships, Deliverables.
3 2999800 Various Appreciation Gifts
4 2999801 Various Appreciation Gifts
3 2999900 Various Products / Equipment and Materials
4 2999901 Various Products / Equipment and Materials
1 3000000 SERVICES
2 3010000 ENGINEERING SERVICES
3 3010100 Project Administration Services
4 3010101 Project Administration Services
3 3010200 Steel / Constr. / Architectural Services
4 3010201 Scale models
4 3010202 Architectural Services
4 3010203 Strength Analyses
3 3010300 Process/Utilities/Piping/HVAC Services
4 3010301 Process Services
4 3010302 Utility Services
4 3010303 Piping Services
4 3010304 HVAC Services
3 3010400 Equipment / Mechanical Services
4 3010401 Mechanical Engineering Services
3 3010500 Electro/Control/Instrumentation Services
4 3010501 Electro Engineering Services
4 3010502 Instrumentation Engineering Services
3 3010600 Pipelines Services
4 3010601 Pipeline Engineering Services
3 3010700 Subsea/Diving/ROV Technology Services
4 3010701 ROV Technology Services
3 3010800 Drilling/Completion Eng. Services.
4 3010801 Drilling Engineering Services

4 3010802 Completion Engineering Services
3 3010900 Civil/Structure Engineering Services
4 3010901 Civil Engineering Services
4 3010902 Concrete Structures Services
3 3011000 Waste Water Treatment Services
3 3011100 Weight Control Services
3 3011200 Tele Communication Services
3 3011300 Material Technology Services
3 3011400 Marine/Hydro/Aerodynamics Services
3 3011500 Fire and Gas Protection System Services
4 3011501 Fire and Gas Protection System Services
3 3011600 Petroleum Engineering Services
3 3011700 Construction Management Services
3 3011800 Mud Engineering Services
3 3019900 Other Engineering Services
2 3020000 CONSULTANCY SERVICES
3 3020100 Quality Assurance QA/QC Consultancy
3 3020200 Safety, Health and Environm. Consultancy
4 3020201 Safety, Health, Environm. Consultancy
3 3020300 Risk Analysis Consultancy
4 3020301 Risk Analyses Consultancy
3 3020500 Training, Personnel Systems Consultancy
4 3020501 Training, Personnel Systems Consultancy
3 3020600 Legal Consultancy
4 3020601 Legal Consultancy
3 3020700 Purchasing and Contract Consultancy
3 3020800 Cost and Planning Consultancy
3 3020900 Material Administration Consultancy
3 3021000 Technical Documentation Consultancy
3 3021100 Advertising / PA / PR Consultancy
4 3021101 Advertising / PA / PR- Consultancy
4 3021102 Social Media Advertising & Agencies
4 3021103 External Job Boards
3 3021200 Marketing and Market Research
3 3021300 Translation Services and Manual Writing
3 3021400 Welding and jointing Consultancy
3 3021500 Warranty Surveyors
4 3021501 Warranty Surveyors
3 3021600 Third Party Evaluation / Verification
4 3021601 Third Party Evaluation Services
4 3021602 Third Party Verification Services
3 3021700 Energy Conservation Consultancy
3 3021800 Simulation Consultancy
4 3021801 Process Simulation Consultancy
3 3021900 Decommissioning and Abandonment
3 3022000 Meteorological Consultancy
3 3022100 Staff Search, Staff Selection
4 3022101 Executive Search and Selection
4 3022102 Staff Search - Key positions
4 3022103 Staff Search - Specialists
4 3022104 Employee Screening
3 3022200 Sub-surface Consultancy
3 3022300 Industrial Design, Web Design etc.
3 3022400 Marine Consultancy
4 3022401 Marine Consultancy
3 3022500 Subsea Consultancy
4 3022501 Subsea Consultancy
3 3022600 Career / Outplacement Consultancy
3 3029800 Management and Business Development
4 3029801 General Management Consultancy
4 3029802 Business Development Consultancy
3 3029900 Other Consultancy Services
2 3030000 RESEARCH & DEVELOPMENT SERVICES (R&D)
3 3030100 Engineering (R&D) Services
4 3030101 Technical Licenses
4 3030102 Design (R&D) Services
3 3030200 Geological and Geophysical (R&D)

3 3030300 Drilling (R&D) Services
3 3030400 Subsea (R&D) Services
4 3030401 Subsea (R&D) Services
3 3030500 Safety and Environmental (R&D) Services
3 3030600 Information Technology (R&D) Services
3 3030700 Oil and Gas Production (R&D) Services
4 3030701 Oil and Gas Production (R&D) Services
3 3030800 Refinery / Downstream Technology (R&D)
3 3030900 Petrochemicals (R&D) Services
3 3031000 Materials (R&D) Services
3 3031100 Reservoir (R & D) Services
3 3039900 Other Research and Development Services
2 3040000 EXPLORATION AND PROD. DRILLING SERVICES
3 3040100 Drilling Rigs
4 3040101 Semisubmersible Drilling Rigs
4 3040102 Jackup Drilling Rigs
3 3040200 Production Drilling Services
3 3040300 Conductor Piling Services
3 3040400 Cementing Services
3 3040500 Casing / Tubing Running Services
3 3040600 Casing Cutting Services
3 3040700 Mud Logging Services
3 3040800 MWD / LWD Services
3 3040900 Down Hole Motors Services
3 3041000 Electrical Logging / Calibration
3 3041100 Coring Services
3 3041200 Fishing / Wireline Fishing Services
3 3041300 Work-over Services
4 3041301 Work-over Services
3 3041400 Snubbing Services
3 3041500 Drilling Fluids/Mud Services
3 3041600 Well Testing Services
3 3041700 Underreaming Services
3 3041800 Exploration and Development Services
3 3041900 Liner Float, Hangers, Running Services
3 3042000 Reservoir Services
3 3042100 Well Completion Services
3 3042200 Well Overhauling/Stimulation Services
3 3042300 Wellhead Services
4 3042301 Wellhead Services
3 3042400 Wireline Services
4 3042401 Wireline Services
3 3042500 Directional Surveying Services
3 3042600 Cutting Injections / Cutting Disposal
4 3042601 Cutting Injections Services
4 3042602 Cutting Disposal Services
3 3042700 Recutting Inspections Services
3 3042800 Rental of Drill Pipe Services
3 3042900 Cased Hole Logging Services
4 3042901 Cased Hole Logging Services
3 3043000 Drill Stem Testing Services
3 3043100 Directional Drilling Services
3 3043200 Perforating Services
4 3043201 Perforating Services
3 3043300 Gravel Packing Services
3 3043400 Fracturating Services
3 3043500 Plugging Services
3 3043600 Well Wash Services
3 3043700 Coiled Tubing Services
3 3043800 Pumping Services
3 3043900 Fluid / Bottom Hole Sampling Services
3 3044000 OCTG Services
4 3044001 OCTG Services
3 3044100 Well Crisis Management Services
3 3049900 Other Drilling Services
2 3050000 CONSTR./MODIF./REPAIR/MAINT./ REFURB.
3 3050100 Steel/Piping/Mechanical Services

4 3050101 Steel/Piping/Mechanical Services
4 3050102 Mechanical Assembly Services
4 3050103 Prototyping Services
4 3050104 Refurbishment Services
4 3050105 Module Fabrication
4 3050106 Threading
3 3050200 Electrical/Instrument Services
4 3050201 Electrical / Instrument Services
3 3050300 Ventilation/Heating/Sanitary Services
4 3050301 Ventilation / Heating Services
4 3050302 Sanitary Services
3 3050400 Waste Disposal / Drainage Services
4 3050401 Waste Disposal Services
4 3050402 Drainage Services
3 3050500 Insulation Services
4 3050501 Insulation Services
3 3050600 Surface Treatment Services
4 3050601 Sandblasting Services
4 3050602 Painting Services
4 3050603 Coating Services
4 3050604 Grinding Services
4 3050605 Bonding Services
3 3050700 Scaffolding/Rigging/Covering Services
4 3050701 Scaffolding/Rigging/Covering Services
3 3050800 Civil Works / Building Services
4 3050801 Civil Works / Building Services
4 3050802 Office Renovation Services
3 3050900 Diving / ROV / Submersible operations
3 3051000 Subsea Construction Services
3 3051100 Diesel Aggregates / Generators Services
3 3051200 Pipe Cutting and Pipe Bending Services
4 3051201 Pipe Cutting
4 3051202 Pipe Bending
4 3051203 Inductive Bending
3 3051300 Industrial Cleaning Services
4 3051301 Industrial Cleaning Services
3 3051400 Catalyst Handling / RegenerationServices
3 3051500 Bolt Tensioning Services
3 3051600 Rope Access Services
4 3051601 Rope Access Services
3 3051700 Welding and Jointing Services
4 3051701 Welding Services
4 3051702 Cladding
3 3051800 Safety/Security/Firefighting System
3 3051900 Preservation of Mech./El. Components
4 3051901 Preservation of Mech./El. Components
3 3052000 Maint./Mod. of Pumps/Rotating Equipment
4 3052001 Maint./Mod. of Pumps/Rotating Equipment
3 3052100 Heat Treating and Demagnetising Services
4 3052101 Heat Treating Services
3 3052200 Vehicle Maintenance and Repair Services
3 3052300 Machining Services
4 3052301 Machining Services
4 3052302 On-Site Machining Services
3 3052400 Site Clean-up Services
3 3052500 Biological Sludge Treatment Services
3 3052600 Tank Bottom Sludge Treatment Services
3 3052700 Piling Services
3 3053000 Valve Management Services
3 3053100 Crane Management Services
3 3059900 Other Construction / Maintenance Services
2 3060000 INST. SERVICES / MARINE CONTRACTING
3 3060100 Pipelaying / Cable Laying Services
3 3060200 Trenching and Excavation Services
3 3060300 Diving / ROV / Submersible operations
4 3060301 Diving / ROV / Submersible operations
3 3060400 Cranes/Crane Barges/Heavy Lift Vessels

4 3060401 Crane Handling Services
3 3060500 Hook-up and Comm Incl. Marine Inst.
4 3060501 Commissioning Services
3 3060600 Mobile Production Units
3 3060700 Dredging Services
3 3060800 Gravel and Rock Dumping Services
3 3060900 Floating Storage Unit (FSU)
3 3061000 Subsea Pipeline Protection Services
3 3061100 Installation of Subsea Packages
4 3061101 Installation of Subsea Packages
3 3061300 Mooring System Services
3 3069900 Other Installation Services
2 3070000 INSPECTION/CONTROL AND TESTING SERVICES
3 3070100 Non-Destructive Testing (NDT) Services
4 3070101 Mechanical Testing Services
3 3070200 Pipeline Services
4 3070201 Pipeline Flushing / Pigging Services
4 3070202 Pipeline External / Internal Inspection
3 3070300 Surface Treatment Inspection Services
3 3070400 Pressure Testing Services
4 3070401 Pressure Testing Services
3 3070500 Instrument Testing/Calibration Services
4 3070501 Instrument Testing/Calibration Services
3 3070600 Load Testing Services
4 3070601 Load Testing Services
3 3070700 Diving / ROV operations
3 3070800 Laboratory Testing Services
4 3070801 Laboratory Testing Services
3 3071000 Dimensional Control / Verification
4 3071001 Stretch Testing Services
4 3071002 Dimensional Control Services
3 3071200 Third Party Measurement Services
4 3071201 Third Party Measurement Services
4 3071202 Third Party Inspection Services
3 3075000 System Integration Testing
4 3075001 System Testing, Land
4 3075002 System Testing, Shallow Water
4 3075003 Extended Factory Acceptance Testing EFAT
3 3079900 Other Testing Services
2 3080000 TRANSPORTATION/ SUPPLY/DISPOSAL SERVICES
3 3080100 Tugs/ROV Support/Diving Support Vessels
4 3080101 Tugboats
4 3080102 ROV Support Vessels
4 3080103 Diving Support Vessels
3 3080200 Barges
3 3080300 Air Transport (Passengers and Freight)
4 3080301 Air Transport, Freight
4 3080302 Air Transport, Passengers
3 3080400 Accomodation Platforms/Vessels
3 3080500 Rental of Containers, Baskets etc
3 3080600 Disposal and Waste Transport Services
4 3080601 Disposal and Waste Transport Services
3 3080700 Salvage Services etc.
3 3080800 Rental of Cranes and Special Vehicles
4 3080801 Rental of Cranes and Special Vehicles
3 3080900 Courier Services
4 3080901 Courier Services
3 3081000 Freight Forwarding/Logistics Management
4 3081001 Freight Forwarding/Logistics Management
3 3081100 Transport Broker and Agent Services
4 3081101 Transport Broker and Agent Services
3 3081200 Supply Base/Warehouse/Storage Services
4 3081201 Warehouse / Storage Services
3 3081300 Ship Chandler Services
3 3081400 Bulk Truck Transportation Services
3 3081500 Truck Transportation Services
3 3081600 Moving Services

4 3081601 Corporate Moving Services
3 3081700 Supply Vessels
4 3081701 Supply Vessel Services
3 3081800 Stand-by Vessels
3 3089900 Other Transportation / Supply and Disposal Services
2 3090000 SURVEYING/POSITIONING SERVICES
3 3090100 Soil Investigation Services
3 3090200 Navigation / Positioning Services
3 3090300 Geotechnical Services
3 3090400 Geophys. and Hydrograph. Site Survey
3 3090500 Oceanographic Services
3 3090600 Rig Positioning Services
3 3090700 Photogrammetry Surveying Services
3 3090800 Chart and Map Production Services
3 3090900 Rental of Surveying/Positioning Equipm.
3 3091000 Survey and Positioning Support Services
4 3091001 Survey and Positioning Support Services
3 3099900 Other Surveying / Positioning Services
2 3100000 SEISMIC/GEOPHYSICAL SERVICES
3 3100100 2D/3D/ 4D Seismic Data Acquisition
4 3100101 2D/3D/4D Seismic Data Acqui.Services
3 3100200 2D / 3D / 4D Seismic Data Processing
3 3100300 Site Surveys Services
3 3100400 Well Velocity Survey and Processing
3 3100500 Rental of Seismic Equipment
3 3100600 Seismic Data Interpretation Services
3 3100700 OBC Seismic Acquisition Services
3 3100800 OBC Seismic Processing Services
3 3100900 OBC Seismic Data Interpretation Services
3 3101000 Electrom. Seismic Acquisition Services
3 3101100 Electrom. Seismic Processing Services
3 3101200 Electrom. Seismic Data Interpretation
3 3101300 Gravity Acquisition and Processing
3 3101400 Magnetic Acquisition and Processing
3 3101500 Speculative Seismic/Geophysical Studies
3 3109900 Other Seismic Services
2 3110000 IT / COMMUNICATION SERVICES
3 3110100 Network Installation/Support Services
4 3110101 Network Installation/Support Services
3 3110200 Software Development and Support
4 3110201 Software Development & Support Services
4 3110202 IT System Testing Services
3 3110300 Computer Based Modelling Services
4 3110301 Computer Based Modelling Services
3 3110400 Computer Based Simul./Training Programs
3 3110500 CAD/CAP Services
3 3110600 Hardware Installation Support Services
4 3110601 Hardware Installation Support Services
3 3110700 Operating Systems Installation / Support
4 3110701 Operating Systems Installation / Support
3 3110800 User Support/Help Desk Services
3 3110900 IT Management Consultancy Services
4 3110901 IT Management Consultancy Services
3 3111000 Data Management Services
3 3111100 General IS / IT Consultancy Services
3 3112000 Telecommunication Installation Services
4 3112001 Telecom Installation Services
3 3112100 Data and Message Transmitting Services
3 3112200 Rental of Telecom Lines
3 3112300 Telecom Subscription Services
4 3112301 Telecom Subscription Services
3 3112400 Public Address System Services
3 3112500 IT / Hardware Support & Maint. Service
4 3112501 Computer Hardware
4 3112502 IT Infrastructure
4 3112503 Computer Peripherals
4 3112504 Mobility & (Tele)Communication

4 3112505 Audio / Video Equipment
3 3119800 Other Communication Services
3 3119900 Other Information Systems / IT Services
2 3120000 PETROLEUM TECHNOLOGY SERVICES
3 3120700 Geophysical Interpretation Services
3 3120800 Petrophysical Interpretation Services
3 3120900 Geological Evaluation Services
4 3120901 Geological Evaluation Services
3 3121000 Reservoir Evaluation Services
3 3129900 Other Petroleum Technology Services
2 3130000 CERTIFICATION AND INTEGRITY SERVICES
3 3130100 Quality Management Systems Certification
4 3130101 Quality Management Systems Certification
3 3130200 Environmental Management Systems Cert.
3 3130300 Safety Management Systems Certification
3 3130400 Info. Security Management System Cert.
3 3130500 Certification of Welders
3 3130600 Certification of NDT-Personnel
3 3130700 Certification of Machinery
3 3130800 Cert. of Cranes and Lifting Appliances
4 3130801 Cert. of Cranes and Lifting Appliances
3 3130900 Certification of Pressurised Equipment
4 3130901 Certification of Pressurised Equipment
3 3131000 Eval. and Cert. of SW and Electronics
3 3131100 Notified Body for Machinery
3 3131200 Notified Body for simple Press. Vessels
3 3131300 Notified Body for Tele Terminal Equipm.
3 3131400 Notified Body for Medical Equipment
3 3131500 Notified Body for Personal Prot. Equipm.
3 3131600 Notified Body for Lifts
3 3131700 Notified Body for Press. Cont. Equipm.
3 3133000 Integrity Management Services
4 3133001 Integrity Management Services
3 3139900 Other Certification Services
2 3140000 FINANCIAL AND INSURANCE SERVICES
3 3140100 Banking Services
4 3140101 Banking Services
3 3140200 Monetary Intermediation Services
3 3140300 Credit Granting Services
4 3140301 Credit Granting Services
3 3140400 Financial Markets Administr. Services
3 3140500 Security Broking and Fund Management
4 3140501 Security Broking and Fund Management
3 3141800 Financial Management Consultancy
3 3141900 Various Banking and Financial Services
4 3141901 Credit Card Services
3 3142000 Accounting Services
4 3142001 Financial Accounting Services
3 3142100 Auditing Services
4 3142101 Financial Auditing Services
3 3143000 Life Insurance Services
4 3143001 Life Insurance Services
3 3143100 Pension Funding Services
4 3143101 Pension Funding Services
3 3143200 Non-Life Insurance Services
4 3143201 Travel Insurance Services
4 3143202 Property Insurance Services
3 3143300 Insurance Broking Services
4 3143301 Insurance Broking Services
3 3143900 Other Insurance Services
2 3150000 DECOMMISSIONING AND ABANDONMENT SERVICES
3 3150100 Project Management Services
3 3150200 Marine Services
3 3150300 Subsea Services
3 3150400 Well Services
3 3150500 Cutting Services
3 3150600 Site Services

4 3150601 Site Services
3 3150700 Waste Management Services
4 3150701 Waste Management Services
3 3150800 Integrity / Legislation Services
3 3159900 Other Decommissioning & Abandon Services
2 3200000 SERVICE STATION MAINTENANCE / SERVICES
3 3200100 Car Wash Equipment Maintenance/Services
3 3200200 Canopy Equipment Maintenance/Services
3 3200300 Maintenance / Service of Signs
3 3200400 Service Station Pumps Services
3 3200500 Payment Terminal Maintenance/Services
3 3200600 Service of Service Station Tanks
3 3209900 Other Service Station Maintenance / Services
2 3500000 SERVICES FOR PROD. OF ALUMINIUM PRODUCTS
3 3500100 Relining of Furnaces
3 3500200 Scrap Conversion
3 3500300 Anodizing
3 3500400 Powder Painting
3 3500500 Mechanical Fabricating
2 3900000 EXCHANGE AND THROUGHPUT SERVICES
3 3900100 EXCHANGE AND THROUGHPUT SERVICES
2 3990000 VARIOUS SUPPORTING SERVICES
3 3990200 Temporary Accomodation / Camps Services
4 3990201 Temporary Accomodation
4 3990202 Long Term Accomodation
3 3990300 Catering Services
4 3990301 Catering Services
4 3990302 Coffee Bar Services
4 3990303 Restaurant Services
4 3990304 Coffee Machines
4 3990305 Water Machines
4 3990306 Vending Machines
4 3990307 Fruit Delivery Services
3 3990400 Cleaning / Laundry Services
4 3990401 Cleaning Services
4 3990402 Laundry Services
3 3990500 Security Services
4 3990501 Security Services
3 3990600 Rental of Equipment
4 3990601 Rental of Equipment
3 3990700 Printing and Copying Services
4 3990701 Printing and Copying Services
4 3990702 Repographics & archive services
3 3990800 Photo / Film / Video Services
3 3990900 Technical Drafting Services
3 3991000 Travel Agencies, Car Rental Services
4 3991001 Travel Agencies
4 3991002 Car Rental
4 3991003 Taxi Services
4 3991004 Various Public Transportation
4 3991005 Tracking of Travel Info / Travellers
3 3991200 Medical Services
4 3991201 Medical Services
3 3991300 Training Course / Seminar Services
4 3991301 Safety Courses / Training
4 3991302 Management Coaching
4 3991303 Leadership Training Courses
4 3991304 General Employee Training Courses
3 3991400 Recreation Services
4 3991401 Recreation Services
3 3991500 Hotel and Conference Services
4 3991501 Hotels
4 3991502 Meetings/Events
3 3991600 Rental of Housing and Office Space etc.
4 3991601 Office Rent
4 3991602 FM Service Charges
4 3991603 Mobile Offices

4 3991604 Car Parking
3 3991700 Energy/Electrical Power Supply Services
4 3991701 Energy/Electrical Power Supply Services
3 3991800 Manpower, Temporary Staff - Technical
4 3991801 Temporary Staff - Blue Collar
4 3991802 Temporary Staff - Subcontractors 03991803
Temporary Staff - Engineers
3 3991900 Manpower, Temporary Staff – Administr.
4 3991901 Temporary Staff - Business Management
4 3991901 Temporary Staff - Commercial / Admin.
4 3991902 Temporary Staff - IT
3 3992000 Library Services
3 3992100 Equipment Brokerage Services
3 3992200 Horticultural / Grounds Maintenance
4 3992201 Property Maintenance
4 3992202 Grounds Maintenance
4 3992203 Plants and Flowers
3 3992300 Water Supply Services
3 3999800 Global Mobility Services
4 3999801 Household Relocation Services
4 3999802 Relocation and Immigration Support
4 3999803 Global Mobility Tax Advisory
4 3999804 Visa Services
3 3999900 Various Supporting Services
4 3999901 Internal Postal Services
1 4000000 MULTI DISC. SERVICES (PACKAGES)
2 4010000 ENGINEERING / PROCUREMENT
3 4010100 Steel/Concrete Structures/Platforms (EP)
3 4010200 Floating Facilities/Systems (EP)
3 4010300 Subsea Systems (EP)
3 4010400 Pipeline Systems (EP)
3 4010500 Risers (EP)
3 4010600 UMBILICALS (EP)
3 4010700 Terminal/Oil Movement Systems (EP)
3 4010800 Accomm./Office/Workshop/Storage Modules (EP)
3 4010900 Process Modules/Packages (EP)
3 4011000 Utilities Modules/Packages (EP)
3 4011100 Drilling Modules/Packages (EP)
3 4011200 Buildings including Service Stations (EP)
3 4019900 Other Systems
2 4020000 ENGINEERING / PROCUREMENT / CONSTRUCTION
3 4020100 Steel/Concrete Structures/Platforms (EPC)
3 4020200 Floating Facilities/Systems (EPC)
3 4020300 Subsea - SWP Top Materials (EPC)
4 4020301 EPC - Subsea Structures
4 4020302 EPC - Wellhead Equipment
4 4020303 EPC - Tooling
4 4020304 EPC - X-mas Tree Systems
4 4020305 EPC - Production Control Systems
4 4020306 EPC - Workover Systems
4 4020307 EPC - Connection Syst. & Pipeline Prod.
4 4020308 EPC - Subsea Power and Boosting
3 4020400 Pipeline Systems (EPC)
3 4020500 Risers (EPC)
3 4020600 UMBILICALS (EPC)
3 4020700 Terminal/Oil Movement Systems (EPC)
3 4020800 Accomm./Office/Workshop/Storage Modules (EPC)
3 4020900 Process Modules/Packages (EPC)
4 4020901 Hydrate Inhibitions Modules/Skid
4 4020902 Water Injections Modules/Skid
4 4020903 Sand and Solids Modules/Skid
4 4020904 Produced Water Modules/Skid
4 4020905 CEC Modules/Skid
4 4020906 Separation Modules/Skid
4 4020907 TEG Contactor and Regenerator
4 4020908 Mercury Removal Modules/Skid
4 4020909 Sulphate Removal Modules/Skid

```
            3 4021000    Utilities Modules/Packages (EPC)
            3 4021100    Drilling Modules/Packages (EPC)
            3 4021200    Buildings incl. Service Stations (EPC)
            3 4029900    Other Systems
      2 4030000    ENG./ PROC./CONSTRUCTION/INSTALLATION
            3 4030100    Steel/Concrete Structures/Platforms (EPCI)
            3 4030200    Floating Facilities/Systems (EPCI)
            3 4030300    Subsea Systems (EPCI)
            3 4030400    Pipeline Systems (EPCI)
            3 4030500    Risers (EPCI)
            3 4030600    UMBILICALS (EPCI)
            3 4030700    Terminal/Oil Movement Systems (EPCI)
            3 4030800    Accomm./Office/Workshop/Storage Modules (EPCI)
            3 4030900    Process Modules/Packages (EPCI)
            3 4031000    Utilities Modules/Packages (EPCI)
            3 4031100    Drilling Modules/Packages (EPCI)
            3 4031200    Buildings incl. Service Stations (EPCI)
            3 4039900    Other Systems
      2 4040000    ENG./PROC./CONSTR./INST./COMMISSIONING
            3 4040100    Steel/Concrete Structures/Platforms (EPCIC)
            3 4040200    Floating Facilities/Systems (EPCIC)
            3 4040300    Subsea Systems (EPCIC)
            3 4040400    Pipeline Systems (EPCIC)
            3 4040500    Risers (EPCIC)
            3 4040600    UMBILICALS (EPCIC)
            3 4040700    Terminal/Oil Movement Systems (EPCIC)
            3 4040800    Accomm./Office/Workshop/Storage Modules (EPCIC)
            3 4040900    Process Modules/Packages (EPCIC)
            3 4041000    Utilities Modules/Packages (EPCIC)
            3 4041100    Drilling Modules/Packages (EPCIC)
            3 4041200    Buildings incl. Service Stations (EPCIC)
            3 4049900    Other Systems
      2 4050000    INTEGRATED SERVICES
            3 4050100    Provision of all Facilities of a Site
            3 4050200    All Operations for a Site/Platf.
            3 4050300    Hydrocarbon Reservoir Services
            3 4050400    Well Test / Early Production Services
            3 4050500    Well Management/Construction Services
            3 4050600    Facility and Security Services at Site
            3 4050700    Field Development
1 6000000    INTERNAL MANUFACTURING
      2 6010000    SEMI FABRICATED COMPONENTS
            3 6010100    Machined Press. Cont./Control Component
                  4 6010101    Surface Treated Components PCC
                  4 6010102    Final Machined Components PCC
                  4 6010103    Cladded Components PCC
                  4 6010104    Pre-cladded Components PCC
            3 6010200    Machined Non Press. Contain./Controlling
                  4 6010201    Surface Treated Components NPCC
                  4 6010202    Final Machined Components NPCC
                  4 6010203    Cladded Components NPCC
                  4 6010204    Precladded Components NPCC
            3 6010300    Semi finished components and structures
                  4 6010301    Semi Finished Bars
                  4 6010302    Semi Finished Blocks / Forgings
                  4 6010303    Semi Finished Structures
                  4 6010304    Semi Finished Plates
                  4 6010305    Semi Finished Sleeves / Pipes
1 7000000    SUB-ASSEMBLIES
      2 7010000    SUBSEA SEMI FINISHED SUB-ASSEMBLIES
            3 7010100    Subsea Structures Sub-Assemblies
                  4 7010101    Sub-Assy - Template (Bottom Foundation)
                  4 7010102    Sub-Assy - Manifold Module (PLEM, FLET)
                  4 7010103    Sub-Assy - Protection Structure
                  4 7010104    Sub-Assy - Flowbase
                  4 7010105    Sub-Assy - Pigging Loop Module
                  4 7010106    Sub-Assy - Temporary Structures
```

3 7010200 Wellhead Equipment Sub-Assemblies
4 7010201 Wellhead Running Tools
4 7010202 Wellhead Guide Bases
3 7010300 Subsea XMT Sub-Assemblies
4 7010301 Sub-Assy - HXT
4 7010302 Sub-Assy - VXT
4 7010303 Sub-Assy - Flow Control Module
4 7010304 Sub-Assy - XMT Tubing Hanger
4 7010305 Sub-Assy - XMT Test Skid
4 7010306 Sub-Assy - XMT Tooling
3 7010400 Production Control System Sub-Assemblies
4 7010401 Sub-Assy - Electrical Power Unit (EPU)
4 7010402 Sub-Assy - Hydraulic Power Unit (HPU)
4 7010403 Sub-Assy - Master Control Station (MCS)
4 7010404 Sub-Assy - Local Operating Station
4 7010405 Sub-Assy - Subsea Control Module SCM
4 7010406 Sub-Assy - SCM Mounting Base
4 7010407 Sub-Assy - Subsea Distribution Unit
4 7010408 Sub-Assy - Hydraulic MQC plate
4 7010409 Sub-Assy - Hydr. Jumper park. Receptacle
4 7010410 Sub-Assy -Electr. Conn. Park. Receptacle
4 7010411 Sub-Assy - Electrical Power Test Unit
4 7010412 Sub-Assy - Hydraulic Power Test Unit
4 7010413 Sub-Assy - Hydr.Jumper Test/Flush. Plate
4 7010414 Sub-Assy - SCMMB Test & Flushing Plate
4 7010415 Sub-Assy - SCM Test Stand
4 7010416 Sub-Assy - Subsea Electronics Module SEM
3 7010500 Workover System Sub-Assemblies
4 7010501 Sub-Assy - Workover Riser System
4 7010502 Sub-Assy - Workover Control System
4 7010503 Sub-Assy - Workover Riser Valve System
4 7010504 Sub-Assy - Emergency Disconnect Package
4 7010505 Sub-Assy - Well Control Package
4 7010506 Sub-Assy - Surface Tree
3 7010600 Intervention & Tooling System Sub-Assy
4 7010601 Sub-Assy - Running Tool System
4 7010602 Sub-Assy - ROV Tooling Package
4 7010603 Sub-Assy - Tie-in Connection System
4 7010604 Sub-Assy - Tooling Handling & Shipping
3 7010700 Sub-Assemblies Connection Systems & Pipeline Products
4 7010701 Sub-Assy - Tie-in Equipment
4 7010702 Sub-Assy - Termination Heads
4 7010703 Sub-Assy - Hubs
4 7010704 Sub-Assy - Clamps and Seal Plates
4 7010705 Sub-Assy - Vertical Connection Sys. VCS
4 7010706 Sub-Assy -Horizontal Connection Sys. HCS
4 7010707 Sub-Assy - Guide & Hinge Over Structure
4 7010708 Sub-Assy - Gas Lift, Choke/Flow Contr.
4 7010709 Sub-Assy - Electrical / Optical Jumpers
4 7010710 Sub-Assy - Connection Tool (RTS / BBRTS)
3 7010800 Subsea Boosting Sub-Assemblies
4 7010801 Subsea Processing Units
4 7010802 High Voltage Power Distribution Units
4 7010803 Low Voltage Power Distribution Units
2 7020000 UMBILICAL SEMI FINISHED SUB-ASSEMBLIES
3 7020100 Umbilical Ancillary Sub-Assemblies
4 7020101 Sub-Assy - Anchor Systems
4 7020102 Sub-Assy - Umbilical Clamps
4 7020103 Sub-Assy - Bend Stiffener Connectors
4 7020104 Sub-Assy - Topside Junction Boxes
4 7020105 Sub-Assy - Repair Splices
3 7020200 Umbilical Components Sub-Assemblies
4 7020201 Sub-Assy - Outer Sheathing
3 7020300 Umbilical Termination Sub-Assemblies
4 7020301 Sub-Assy - Topside Termination Template
4 7020302 Sub-Assy - Subsea Termination Interface
4 7020303 Sub-Assy - Pull-in & hang-off Structure

4 7020304 Sub-Assy - Termination Lifting Arrangem.
4 7020305 Sub-Assy - Bend Restrictors / Clamps
4 7020306 Sub-Assy - Transition Joints
4 7020307 Sub-Assy - Subsea Umbilical Termination
2 7030000 MMO SEMI FINISHED SUB-ASSEMBLIES
3 7030100 MMO SEMI FINISHED SUB-ASSEMBLIES
2 7040000 AET SEMI FINISHED SUB-ASSEMBLIES
3 7040100 AET SEMI FINISHED SUB-ASSEMBLIES
1 8000000 FINISHED TOP ASSEMBLIES
2 8010000 SUBSEA TOP ASSEMBLIES
3 8010100 Subsea Structures
4 8010101 Templates (Bottom Foundation Structure)
4 8010102 Manifold Module (PLEM, FLET Module)
4 8010103 Protection Structures
4 8010104 Flowbase
4 8010105 Pigging Loop Modules
4 8010106 Temporary Structures
4 8010107 Guide Posts
4 8010108 Guide Bases
4 8010109 Mud Mats
3 8010200 Wellhead Equipment
4 8010201 Wellheads and Casing Hangers
4 8010202 Redress Spares
3 8010300 Subsea Tooling
4 8010301 Trees Tooling
4 8010302 Wellheads Tooling
4 8010303 ROV Tooling
4 8010304 Workover Tooling
4 8010305 Tie-in Tooling
4 8010306 Test & Flush Tooling
4 8010307 Down Hole Tooling
3 8010400 Horisontal XT
4 8010401 HXT - Production
4 8010402 HXT - Water Injection
4 8010403 HXT - Gas Injection
4 8010404 HXT - Tubing Hanger
4 8010405 HXT - Internal Tree Cap (ITC)
4 8010406 HXT - Tubing Hanger Plugs
4 8010407 HXT - Blowout Preventer (BOP)
3 8010500 Vertical XT
4 8000501 VXT - Production
4 8000502 VXT - Water Injection
4 8000503 VXT - Gas Injection
4 8000504 VXT - Tubing Hanger
4 8000505 VXT - Plugs
4 8000506 VXT - Blowout Preventer (BOP)
3 8010600 Flow Control Module
4 8010601 Flow Control Module - Production
4 8010602 Flow Control Module - Water Injection
4 8010603 Flow Control Module - Gas Injection
3 8010700 Production Control System
4 8010701 Electrical Power Unit (EPU)
4 8010702 Hydraulic Power Unit (HPU)
4 8010703 Master Control Station (MCS)
4 8010704 Local Operating Station
4 8010705 Subsea Control Module (SCM)
4 8010706 Subsea Control Module Mounting Base
4 8010707 Subsea Distribution Unit
4 8010708 Hydraulic MQC plate
4 8010709 Hydraulic Jumper Parking Receptacles
4 8010710 El./Opt. Connector parking receptacle
4 8010711 Electrical Power Test Unit
4 8010712 Hydraulic Power Test Unit
4 8010713 Hydraulic Jumper Test & Flushing Plate
4 8010714 Electronic Test Unit
4 8010715 SCMMB Test & Flushing Plate
4 8010716 Subsea Control Module Test Stand

4 8010717 Controls Containers
4 8010718 Subsea Electronics Module (SEM)
3 8010800 Workover System
4 8010801 Workover Riser System
4 8010802 Workover Control System
4 8010803 Workover Riser Valve System
4 8010804 Emergency Disconnect Package
4 8010805 Well Control Package
4 8010806 Surface Tree
4 8010807 Umbilical Reel
4 8010808 Landing String
4 8010809 Well Intervention Equipment
3 8010900 Connection Systems & Pipeline Products
4 8010901 Termination Heads
4 8010902 Hubs
4 8010903 Clamps and Seal Plates
4 8010904 Vertical Connection System
4 8010905 Horizontal Connection System
4 8010906 Guide & Hinge Over Structure
4 8010907 Tie-In, Gas Lift, Choke/Flow Contr. Mod.
3 8011000 Subsea Boosting
4 8011001 Subsea MultiBooster
4 8011002 Subsea HybridBooster
4 8011003 Subsea LiquidBooster
2 8020000 UMBILICAL TOP ASSEMBLIES
3 8020100 Umbilical Ancillary
4 8020101 Anchor Systems
4 8020102 Umbilical Clamps
4 8020103 Bend Stiiffener Connectors
4 8020104 Topside junction Boxs
4 8020105 Umbilical Repair kits
4 8020106 Umbilical Repair Splices
3 8020200 Umbilical Components
4 8020201 Dynamic X-sections
4 8020202 Electric Cables
4 8020203 Fillers
4 8020204 Fluid
4 8020205 Fiber Optic Cables
4 8020206 PVC Conduits
4 8020207 Outer Sheeting
4 8020208 Static X-section
4 8020209 Tubex
4 8020210 Weight Element
3 8020300 Umbilical Termination
4 8020301 Topside Termination Templates
4 8020302 Subsea Termination Interfaces
4 8020303 Pull-in and hang-offs
4 8020304 Termination lifting Arrangements
4 8020305 Bend Restrictors/Clamps
4 8020306 Transition Joints
4 8020307 Subsea Umbilical Termination
3 8020400 Umbilical Sub Supply
4 8020401 Transportation Reels
4 8020402 J-tube Seals
4 8020403 Bend Stiffeners
4 8020404 Buoyancy Elements
2 8030000 MMO TOP ASSEMBLIES
3 8030100 Gas Lift Top Assemblies
3 8030200 Platform Tie-in Top Assemblies
3 8030300 Platform Drilling Top Assemblies
3 8030400 Platform Modification Top Assemblies
2 8040000 AET TOP ASSEMBLIES
3 8040100 Offshore oil & gas production facilities
4 8040101 Oil & gas treatment
4 8040102 Oil & gas storage, offloading & export
4 8040103 Utility and process support systems
4 8040104 Drilling facilities

4 8040105 Living quarters
3 8040200 Floaters (Semis, TLPs, SPARs, FPSOs)
4 8040201 Semi
4 8040202 TLP
4 8040203 SPAR
4 8040204 FPSO
3 8040300 Riser Systems
4 8040301 Deepwater Risers
4 8040302 Integrated platform/riser/mooring
4 8040303 Integrated platform/riser/mooring
1 9000000 SYSTEMS AND PACKAGES
2 9010000 CLIENT - AKOFS
3 9010100 FIELD - AKOFS SESV I
4 9010101 SYSTEM - AKOFS SESV I
3 9010200 FIELD - AKOFS SESV II
4 9010201 SYSTEM - AKOFS SESV II
2 9020000 CLIENT - BG GROUP
3 9020100 FIELD - EVEREST
4 9020101 SYSTEM - EVEREST
2 9030000 CLIENT - BHP Billiton
3 9030100 FIELD - KIETH
4 9030101 SYSTEM - KIETH
2 9040000 CLIENT - BP
3 9040100 FIELD - ANDREW
4 9040101 SYSTEM - ANDREW
3 9040200 FIELD - BRUCE
4 9040201 SYSTEM - BRUCE
3 9040300 FIELD - CYRUS
4 9040301 SYSTEM - CYRUS
4 9040302 SYSTEM - CYRUS C3
4 9040303 SYSTEM - CYRUS C4
3 9040300 FIELD - ETAP
4 9040301 SYSTEM -EGRET
4 9040302 SYSTEM - HERON
4 9040303 SYSTEM - MACHAR
4 9040304 SYSTEM - MADOES
4 9040305 SYSTEM - MARNOCK
4 9040306 SYSTEM - MIRREN
4 9040307 SYSTEM - MONAN
4 9040308 SYSTEM - MUNGO
4 9040309 SYSTEM - SKUA
3 9040400 FIELD - EVEREST
4 9040401 SYSTEM - EGRET
4 9040402 SYSTEM - HERON
3 9040500 FIELD - NAM CON SON
4 9040501 SYSTEM - NAM CON SON
3 9040600 FIELD - NLGP
4 9040601 SYSTEM - NLGP
3 9040700 FIELD - SCHIEHALLION
4 9040701 SYSTEM - SCHIEHALLION
3 9040800 FIELD - SULLOM VOE
4 9040801 SYSTEM - SULLOM VOE
2 9050000 CLIENT - BW GROUP
2 9060000 CLIENT - CHEVRON
3 9060100 FIELD - JACK AND ST MALO
4 9060101 SYSTEM - JACK AND ST MALO
2 9070000 CLIENT - CNOOC
3 9070100 FIELD - YA CHENG
4 9070101 SYSTEM - YA CHENG
3 9070200 FIELD - LIU HUA
4 9070201 SYSTEM - LIU HUA
2 9080000 CLIENT - CONOCOPHILLIPS
2 9090000 CLIENT - DET NORSKE OLJESELSKAPET
3 9090100 FIELD - ALVHEIM
4 9090101 SYSTEM - ALVHEIM
3 9090200 FIELD - VOLUND
4 9090201 SYSTEM - VOLUND

3 9090300 FIELD - BØYLA
4 9090301 SYSTEM - BØYLA
2 9100000 CLIENT - DNO
3 9100100 FIELD - BROOM
4 9100101 SYSTEM - BROOM
3 9100200 FIELD - HEATHER
4 9100201 SYSTEM - HEATHER
2 9110000 CLIENT - DONG ENERGY
3 9110100 FIELD - TRYM
4 9110101 SYSTEM - TRYM
3 9110200 FIELD - OSELVAR
4 9110201 SYSTEM - OSELVAR
2 9120000 CLIENT - ENI
3 9120100 FIELD - GOLIAT
4 9120101 SYSTEM - GOLIAT
3 9120200 "FIELD - LONGHORN
4 9120201 SYSTEM - LONGHORN
3 9120300 "FIELD - OYO
4 9120301 SYSTEM - OYO
2 9130000 CLIENT - ENQUEST
3 9130100 FIELD - KRAKEN
4 9130101 SYSTEM - KRAKEN
2 9140000 CLIENT - EXXON MOBILE
2 9150000 CLIENT - FREEPORT-McMoRan
3 9150100 FIELD - KING (Gulf of Mexico)
4 9150101 SYSTEM - KING (Gulf of Mexico)
2 9160000 CLIENT - HARDY OIL&GAS
3 9160100 FIELD - FOR HARDY
4 9160101 SYSTEM - FOR HARDY
2 9170000 CLIENT - HESS
3 9170100 FIELD - GB 201
4 9170101 SYSTEM - GB 201
3 9170200 FIELD - NORTHWESTERN
4 9170201 SYSTEM - NORTH WESTERN
3 9170300 FIELD - HESS GB 114
4 9170301 SYSTEM - GB 114
2 9180000 CLIENT - ITHACA ENERGY
2 9190000 CLIENT - LUNDIN
3 9190100 FIELD - BROOM
4 9190101 SYSTEM - BROOM
3 9190200 FIELD - BRYNHILD
4 9190201 SYSTEM - BRYNHILD
2 9200000 CLIENT - MAERSK
3 9200100 FIELD - AFFLECK
4 9200101 SYSTEM - AFFLECK
3 9200200 FIELD - DUMBARTON
4 9200201 SYSTEM - DUMBARTON
3 9200300 FIELD - GRYPHON
4 9200301 SYSTEM - GRYPHON
3 9200300 FIELD - MACLURE
4 9200301 SYSTEM - MACLURE
2 9210000 CLIENT - MURPHY
3 9210100 FIELD - KIKEH
4 9210101 SYSTEM - KIKEH
3 9210200 FIELD - MEDUSA
4 9210201 SYSTEM - MEDUSA
2 9220000 CLIENT - NEWFIELD
3 9220100 FIELD - GB161-4
4 9220101 SYSTEM - GB161-4
2 9230000 CLIENT - ORIGIN ENERGY
3 9230100 FIELD - OTWAY
4 9230101 SYSTEM - OTWAY
2 9240000 CLIENT - PETROBRAS
3 9240100 FIELD - 30 XMT 2005
4 9240101 SYSTEM - 30 XMT 2005
3 9240200 FIELD - 45 XMT 2007
4 9240201 SYSTEM - 45 XMT 2007

```
        3 9240300    FIELD - GUARA & LULA
              4 9240301    SYSTEM - GUARA & LULA
        3 9240400    FIELD - LULA (TUPI)
              4 9240401    SYSTEM - LULA (TUPI)
        3 9240500    FIELD - 2011 ANM FA
              4 9240501    SYSTEM - 2011 ANM FA
        3 9240600    FIELD - SLNE Fase II
              4 9240601    SYSTEM - SLNE Fase II
  2 9250000    CLIENT - PETRONAS
        3 9250100    FIELD - KANOWIT
              4 9250101    SYSTEM - KANOWIT
  2 9260000    CLIENT - PREMIER OIL
        3 9260100    FIELD - BREAM
              4 9260101    SYSTEM - BREAM
  2 9270000    CLIENT - PVEP
        3 9270100    FIELD - DAI HUNG
              4 9270101    SYSTEM - DAI HUNG
  2 9280000    CLIENT - RAPID SOLUTION
        3 9280100    FIELD - DEEP WATER
              4 9280101    SYSTEM - DEEP WATER
        3 9280200    FIELD - SHALLOW WATER
              4 9280201    SYSTEM - SHALLOW WATER
  2 9290000    CLIENT - RELIANCE
        3 9290100    FIELD - KG-D6
              4 9290101    SYSTEM - KG-D6
        3 9290200    FIELD - MA-D6
              4 9290201    SYSTEM - MA-D6
  2 9300000    CLIENT - REPSOL
        3 9300100    FIELD - YAMAGUA
              4 9300101    SYSTEM - YAMAGUA
        3 9300200    FIELD - MONTANAZO-LUBINA_PCS
              4 9300201    SYSTEM - MONTANAZO-LUBINA_PCS
  2 9310000    CLIENT - SHELL
        3 9310100    FIELD - BITTERN
              4 9310101    SYSTEM - BITTERN
        3 9310200    FIELD - BRENT
              4 9310201    SYSTEM - BRENT
        3 9310300    FIELD - CORMORANT
              4 9310301    SYSTEM - CORMORANT
        3 9310400    FIELD - CURLEW
              4 9310401    SYSTEM - CURLEW
        3 9310500    FIELD - DRAUGEN
              4 9310501    SYSTEM - DRAUGEN
        3 9310600    FIELD - FULMAR
              4 9310601    SYSTEM - FULMAR
        3 9310700    FIELD - FRAM
              4 9310701    SYSTEM - FRAM
        3 9310800    FIELD - GANNET
              4 9310801    SYSTEM - GANNET
        3 9310900    FIELD - GARN WEST
              4 9310901    SYSTEM - GARN WEST
        3 9311000    FIELD - GUILLEMOT
              4 9311001    SYSTEM - GUILLEMOT
        3 9311100    FIELD - HOWE
              4 9311101    SYSTEM - HOWE
        3 9311200    FIELD - HERON
              4 9311201    SYSTEM - HERON
        3 9311300    FIELD - KESTREL
              4 9311301    SYSTEM - KESTREL
        3 9311400    FIELD - KINGFISHER
              4 9311401    SYSTEM - KINGFISHER
        3 9311500    FIELD - LEMAN
              4 9311501    SYSTEM - LEMAN
        3 9311600    FIELD - MALLARD
              4 9311601    SYSTEM - MALLARD
        3 9311700    FIELD - MANDARIN
              4 9311701    SYSTEM - MANDARIN
```

3 9311800 FIELD - MENSA
4 9311801 SYSTEM - MENSA
3 9311900 FIELD - MERGANSER
4 9311901 SYSTEM - MERGANSER
3 9312000 FIELD - NELSON
4 9312001 SYSTEM - NELSON
3 9312100 FIELD - ORMEN LANGE
4 9312101 SYSTEM - ORMEN LANGE
3 9312200 FIELD - OSPREY/MERLIN FAIRFIELD
4 9312201 SYSTEM - OSPREY/MERLIN FAIRFIELD
3 9312300 FIELD - OSPREY
4 9312301 SYSTEM - OSPREY
3 9312400 FIELD - PENGUIN
4 9312401 SYSTEM - PENGUIN
3 9312500 FIELD - PELICAN
4 9312501 SYSTEM - PELICAN
3 9312600 FIELD - ROGN SOUTH
4 9312601 SYSTEM - ROGN SOUTH
3 9312700 FIELD - SCOTER
4 9312701 SYSTEM - SCOTER
3 9312800 FIELD - STARLING
4 9312801 SYSTEM - STARLING
3 9312900 FIELD - STOCK
4 9312901 SYSTEM - STOCK
3 9313000 FIELD - TEAL
4 9313001 SYSTEM - TEAL
3 9313000 FIELD - TERN
4 9313001 SYSTEM - TERN
2 9320000 CLIENT - SOCAR
2 9330000 CLIENT - SONANGOL
3 9330100 FIELD - GIMBOA
4 9330101 SYSTEM - GIMBOA
2 9340000 CLIENT - STATOIL
3 9340100 FIELD - AASGARD
4 9340101 SYSTEM - AASGARD
3 9340200 FIELD - AASTA HANSTEEN
4 9340201 SYSTEM - AASTA HANSTEEN
3 9340300 FIELD - DIACS
4 9340301 SYSTEM - DIACS
3 9340400 FIELD - FRAM OST
4 9340401 SYSTEM - FRAM OST
3 9340500 FIELD - JOHAN CASTBERG
4 9340501 SYSTEM - JOHAN CASTBERG
3 9340600 FIELD - JOHAN SVERDRUP
4 9340601 SYSTEM - JOHAN SVERDRUP
3 9340700 FIELD - KRAFLA
4 9340701 SYSTEM - KRAFLA
3 9340800 FIELD - KRISTIN
4 9340801 SYSTEM - KRISTIN
3 9340900 FIELD - LIS II
4 9340901 SYSTEM - LIS II
3 9341000 FIELD - MORVIN
4 9341001 SYSTEM - MORVIN
3 9341100 FIELD - NJORD
4 9341101 SYSTEM - NJORD
3 9341200 FIELD - SKULD
4 9341201 SYSTEM - SKULD
3 9341300 FIELD - SVALIN
4 9341301 SYSTEM - SVALIN
3 9341400 FIELD - TOFTE
4 9341401 SYSTEM - TOFTE
3 9341500 FIELD - TROLL
4 9341501 SYSTEM - TROLL
3 9341600 FIELD - TUNE PHASE 1
4 9341601 SYSTEM - TUNE PHASE 1
3 9341700 FIELD - TUNE PHASE 2
4 9341701 SYSTEM - TUNE PHASE 2

```
      3 9341800    FIELD - TYRIHANS
            4 9341801    SYSTEM - TYRIHANS
      3 9341900    FIELD - VIGDIS
            4 9341901    SYSTEM - VIGDIS
      3 9342000    FIELD - VILJE
            4 9342001    SYSTEM - VILJE
      3 9342100    FIELD - VILJE SOER
            4 9342101    SYSTEM - VILJE SOER
      3 9342200    FIELD - VISUND
            4 9342201    SYSTEM - VISUND
      3 9342300    FIELD - VISUND NORD
            4 9342301    SYSTEM - VISUND NORD
2 9350000    CLIENT - TALISMAN
      3 9350100    FIELD - CO4
            4 9350101    SYSTEM - CO4
      3 9350200    FIELD - BLANE
            4 9350201    SYSTEM - BLANE
      3 9350300    FIELD - ENOCH
            4 9350301    SYSTEM - ENOCH
      3 9350400    FIELD - WAGE
            4 9350401    SYSTEM - WAGE
2 9360000    CLIENT - TOTAL
      3 9360100    FIELD - ATLA
            4 9360101    SYSTEM - ATLA
      3 9360200    FIELD - DALIA
            4 9360201    SYSTEM - DALIA PHASE 1
            4 9360202    SYSTEM - DALIA PHASE 1A
            4 9360203    SYSTEM - DALIA PHASE 2
            4 9360204    SYSTEM - DALIA PHASE 3
            4 9360205    SYSTEM - DALIA PHASE 4
      3 9360300    FIELD - HILD
            4 9360301    SYSTEM - HILD
      3 9360400    FIELD - KAOMBO
            4 9360401    SYSTEM - KAOMBO
      3 9360500    FIELD - LILLE FRIGG
            4 9360501    SYSTEM - LILLE FRIGG
      3 9360600    FIELD - MOHO NORD
            4 9360601    SYSTEM - MOHO NORD
2 9370000    CLIENT - VENTURE (CENTRICA)
      3 9370100    FIELD - AMANDA AGATH 2
            4 9370101    SYSTEM - AMANDA AGATH 2
      3 9370200    FIELD - ANN A4 INFILL
            4 9370201    SYSTEM - ANN A4 INFILL
      3 9370300    FIELD - ANNABEL
            4 9370301    SYSTEM - ANNABEL
            4 9370302    SYSTEM - ANNABEL EAST
      3 9370400    FIELD - CHANNON & BARB
            4 9370401    SYSTEM - CHANNON & BARB
      3 9370500    FIELD - CHESNUT
            4 9370501    SYSTEM - CHESNUT
      3 9370600    FIELD - GADWELL
            4 9370601    SYSTEM - GADWELL
      3 9370700    FIELD - GOOSANDER
            4 9370701    SYSTEM - GOOSANDER
      3 9370800    FIELD - GROUSE
            4 9370801    SYSTEM - GROUSE
      3 9370900    FIELD - MALLARD EAST
            4 9370901    SYSTEM - MALLARD EAST
      3 9371000    FIELD - SNS TREE 6 & 7
            4 9371001    SYSTEM - SNS TREE 6 & 7
      3 9371100    FIELD - W1 WELL GOOSAN
            4 9371101    SYSTEM - W1 WELL GOOSAN
2 9380000    CLIENT - WOODSIDE
      3 9380100    FIELD - LAMINARIA
            4 9380101    SYSTEM - LAMINARIA
2 9390000    CLIENT - GDF SUEZ
2 9400000    CLIENT - GAZPROM
```

1 10000000 LIFE CYCLE SERVICE PACKAGES
2 10010000 CLIENT
3 1010100 FIELD
4 1010101 SYSTEM