Management of heterogeneous information for integrated design of multidisciplinary systems

Benjamin GUERINEAU\textsuperscript{ab*}, Chen ZHENG\textsuperscript{a}, Matthieu BRICOGNE\textsuperscript{a}, Alexandre DURUPT\textsuperscript{a}, Louis RIVEST\textsuperscript{b}, Harvey ROWSON\textsuperscript{a}, Benoît EYNARD\textsuperscript{a}

\textsuperscript{a}Laboratoire Roberval, UMR UTC/CNRS 7337, Université de Technologie de Compiègne, Rue du Dr Schweitzer, Compiègne 60203 Cedex, France
\textsuperscript{b}Laboratoire Numérique, Ecole de Technologie Supérieure, 1100 Rue Notre-Dame O, Montréal QC H3C 1K3, Canada

* Corresponding author. Tel.: +33 3 44 23 73 58; fax: +0-000-000-0000. Email address: benjamin.guerineau@utc.fr

Abstract

Multidisciplinary systems (such as Mechatronics or Cyber Physical Systems) are considered as the resulting integration of design expertise from several disciplines such as electrical/electronic, mechanical and computer sciences. As a result, a large number of design data, such as software code, CAD models, 0D/1D and 2D/3D CAE results, etc. are generated by designers and heterogeneous computer-based tools throughout the whole development process. Therefore, effective exchange between designers from different disciplines is required in order to achieve multidisciplinary integration. In order to ensure knowledge sharing between the designers issued from the different disciplines, the heterogeneous information from previous design projects could be captured, elucidated and managed by designers. In this paper, after presenting the multidisciplinary integration during the system design, the importance of effective exchange between designers from different disciplines is highlighted. Then, the existing techniques related to the capture and management of heterogeneous information are presented. Afterwards, an approach helping designers to capture the design data issued from CAD models and 0D/1D and 2D/3D CAE results, is introduced. Finally, the conclusion is drawn and future work is pointed out.

Keywords: Knowledge Based Engineering; Collaborative Design; Data extraction; Mechatronics; Cyber-Physical-Systems; Routine Design

1. Introduction

This paper gives an overview to a knowledge-based engineering (KBE) approach developed for an ongoing research project, focusing on multidisciplinary engineering knowledge capitalization. The information captured from past projects is then stored in an agnostic knowledge solution based on parameters and rules. This work aims to present the extraction process and the semantic issues related to the data processing step.

The following section presents the context of our work. A part of the research work is focused on the design of multidisciplinary systems such as Mechatronic and Cyber-Physical-Systems products in a routine design context [30]. This section also presents the objectives and the problem statement.

The third section investigate the state-of-the-art of heterogeneous data and highlight the fact that, up to present, most studies were focused on the extraction of a domain specific data. Then some techniques are discussed to solve semantic problems in a specific area and is enlarged to multidisciplinary concerns.

Next, the proposition on two different aspects is envisioned. The first aspect is on the identification of six cases of semantic problems and their characterization. The second is our core proposition represented by a general process of extracting and processing heterogeneous data from various sources: Computer Aided Design (CAD) or Digital Mock-Up (DMU), 0D/1D and 2D/3D simulations which are stored into Information Systems (IS) called “third-parties systems”.

Finally, a conclusion summarize our proposition and future work for our research project is exposed. An acknowledgement
will thanks our industrial and academic partners for their respective contribution.

2. Context

Our work is focusing on developing an approach for an ongoing research project called MIMe including academic and industrial partners. This acronym stands for “Module d’Intégration et de Simulation pour la Mécatronique” (Simulation and Integration Module for Mechatronics). As specified in the acronym, this work focuses on mechatronics. Mechatronic products can be defined as “the result of combining the engineering disciplines mechanics, electrics, electronics and IT”. This requires coordinated trans-sectoral cooperation from the people developing the product as well as from the organisational unit [1].

Although, to enlarge the scope of our work, Cyber-Physical-Systems (CPS) are also considered. CPS are defined as “physical and engineered systems whose operations are monitored, coordinated, controlled, and integrated by a computing and communication core” [23].

Mechatronics and CPS are two classes of what are called “multidisciplinary products” [11,32] whose main characteristic is to be developed across different disciplines. These classes of products are becoming more and more complex to develop [31]. Collaboration and integration are unavoidable points in this area and attract attention of both academia and industry [12].

Modern industry is facing ever increasing challenges of more complex products and even sometimes geographically dispersed design teams. It is likely that different design teams use different terms to represent the same information according to their own expertise and discipline, which yields confusion during the collaborative design activities of such complex products. For example, one design team use the unit “degree” to describe the angle while another use “radian”. Specifically, this multidisciplinary context involves the use of various Computer Aided Engineering (CAE) software - at least one or more disciplines and their associated models. These models often contain engineering data in a heterogeneous form that our work aim to extract, process and pour into a knowledge data base for a further use. Heterogeneity is used to characterize data that originate from different sources (CAD, Computer-aided manufacturing, Simulation results, technical metadata, etc.), that could be from different types (physical magnitudes, texts, images, videos, etc.) or by their format.

The concern of a further use is mainly applicable in a routine design context. In a routine design context [30], as opposed to an innovative design context, each project has similarities with previous ones. A capitalization from previous data is thus a possible manner to reduce development time, cost [24] and risks. As a consequence, a KBE approach is a potential solution to capitalize data from past project. Although, most KBE solutions are only structured data based those that are sometimes integrated in a PLM. This article will propose an approach to support extraction process to complete a knowledge data-base from heterogeneous data.

In the context, these heterogeneous data will be extracted from different software, different disciplines and different teams or companies. As an observation, the technical lexicon could change between companies, teams, software, disciplines, etc. Besides the complexity of the multidisciplinary product development, the way to extract data also leads to semantic issues such as misunderstandings. This statement could be a real problem in a collaborative work. The use of different words to characterize the same object for example can impede the collaboration and make it unclear. In most cases humans are able to recognize semantic problems, but it is envisioned to automatize this process through a semantic dictionary.

Our objective is to allow companies that mainly work in a routine design approach to capitalize data from their past projects, focusing first on CAD/DMU, 0D/1D simulations and 2D and 3D simulations. These data are part of the third-parties systems. To achieve this objective, a bridge, represented by our process, will be built between these third-parties systems and the agnostic knowledge base. Agnostic is defined as an adjective “denoting or relating to hardware or software that is compatible with many types of platform or operating system” [22]. The multidisciplinary context and the data heterogeneity from many sources entail to be agnostic and is as a consequence wanted to be free from a specific software editor.

The second objective is to define a semantic dictionary to identify and converge technical terms to improve and ease multidisciplinary collaboration. This second objective support the first exposed above. Next section presents the related work which can partially achieve the two objectives.

3. State-of-the-art

The first objective of the research work is to help designers to extract data sources stored by third-parties systems. The research context presented in Section 2 reveals that the various third-parties systems have been involved during the design process of the so-called multidisciplinary products. Increasing number of data sources are generated accordingly. Therefore it is significant to develop a bridge which can connect the third-parties systems and the knowledge base so that designers can extract the heterogeneous data from the diverse third-parties systems and store them in the KBE system.

KBE is a technology based on dedicated software tools, which is able to capture and systematically reuse product and process engineering knowledge, with the final goal of reducing time and costs of product development [25]. How to extract the heterogeneous data sources and store them in the knowledge base is a subject of increasingly interest both of both academia and industry. Extraction of heterogeneous data sources to support collaborative activities has been studied for long time in the discipline of computer science. According to Batini et al, name heterogeneity is the most common type of semantic heterogeneity [2]. Name heterogeneity refers to the conflicts of terms used to describe different data source. Castano and Antonellis propose a global view that can provide a unified representation of the information in the different sources [4]. Rouset and Reynaud develop an information integration system which provides a uniform query interface to collection of distributed and heterogeneous information sources [26]. In their proposition, the data of the World-Wide Web which are
described by XML documents can be automatically extracted and transferred into a uniform tree model.

Besides the discipline of computer science, much more attention has been paid to the extraction of heterogeneous data generated related to mechanical design. LeBlanc and Fadel propose an Object-Oriented Product Model (OOPM) to extract and store the geometry and surface data [17]. However, the OOPM only focuses on the geometric data. Graening et al reveal the importance of CAE results. They propose methods that allow the extraction of heterogeneous design data of aerodynamic shape which are represented by discrete unstructured surface meshes [8]. The previous review shows that most of existing studies on the extraction of heterogeneous data sources focus on a single discipline (i.e., computer science or mechanics). However, little attention has been paid to the extraction of heterogeneous data generated across different disciplines.

The second objective of our research is to define a semantic dictionary to help designers to identify technical terms and share information. During the design process of multidisciplinary products, different technical terms may be used by designers from different disciplines. Moreover, same information in different design tasks (i.e., design phase) may be represented by different terms. In order to avoid the confusion by misuse of information shared by different design teams during the collaborative design activities, in our semantic context, “meaning, “term” and “object” have to be defined. To define these words and their interactions, the semiotic triangle is introduced.

The semiotic triangle (also known as triangle of reference and triangle of meaning) is a model of how linguistic symbols are related to the objects they represent. This triangle was proposed by Ogden and Richards in The Meaning of Meaning in 1923 [21]. This semiotic triangle is currently the most referred to in the informatics literature on natural processing: meaning results of a relation between a term and an object.

![Fig. 1. Semiotic triangle, adapted from [21]](image)

Fig.1 indicates that when we experience a certain object in association with a certain term, then memory traces are laid down in our brains in virtue of which the mere appearance of the same term in the future will “evok” a “meaning” directed towards this object through the reactivation of impressions stored in memory.

The two solid edges of the triangle are intended to represent what are held to be causal relations of term and meaning on the part of a symbol-using subject [27]. The dashed edge between term and object indicates have no direct and immediate relation to the objects for which they stand. In fact, a relation is only indirectly possible through terms that refer to the objects and the meanings which are activated by the appearance of terms [29]. Therefore our proposition related to the semantic problems focuses on the terminological relationships between the term and the meaning, which will be discussed in the following section.

In this section, regarding the two objective of the research work related to the extraction of heterogeneous data sources and the semantic dictionary, we have reviewed the existing studies on extraction of heterogeneous data and ontological approaches respectively. The semiotic triangle has been introduced to define specific words as well. Next section will present our propositions by considering the limitations of existing studies.

4. Proposition

The proposition is broken down into two parts. The first part support our vision of the semantic problem with six main cases that have been identified. The second part is related to the process from the extraction to the discharge into the KBE data, through the data processing. These two sub-proposition are merged in the general proposed process.

4.1. Semantic dictionary

The first proposition is related to the semantic referential where six main cases have been identified to represent two main situations that could be encountered in a collaborative work. These two main situation are called “synonymy” and “homonymy”. The six cases stand for the different situations that could occur during a project. Before presenting the core of the proposition, synonymy and homonymy will be defined as they are the basement of the six cases.

Synonymy, as described in previous section, defines the terms that are interchangeable in a particular context. Therefore it can help the designers to represent the same or similar information which may be described by different terms. In order to achieve the collaboration and improve the interoperability among the designers with different expertise background, it is necessary to define the relationship of synonymy in the semantic dictionary.

In linguistics, homonym is defined as the relationship of two words if they are pronounced or spelled the same way but have different meanings. Such relationship also exists in the engineering disciplines. The design of mechatronic systems requires the multidisciplinary collaboration. However, the same term may signify different information in different disciplines. The semantic dictionary should provide a guide to the designers by which they can distinguish the different information which may be assigned by the same term to avoid the confusion during the design process.

Hyponymy and hypernymy also need to be presented for the proposition explanations. Hyponymy and hypernymy are a relation of granularity between two terms. In our research context, same or similar information during the different design
phase may be represented by different terms with different abstraction levels.

Then, the term “inclusion” is proposed to define the relation of a shared meaning between two spaces. This can often be the case in a collaborative work between two disciplines.

Our proposition is based on these concepts and the six cases are graphically represented on Fig. 2. Three cases deals with synonymy and the three others with homonymy. Each of the three cases is categorized as “Granular”, “Partial” or “Total”. Ti and M, respectively stands for “term” and “meaning”.

The granular-synonymy consist of merging synonymy and the concept of hyponymy and hypernymy by including different granularities. The term T1 is linked to the meaning M1 with is more precise than the term T2 and the associated meaning M2. A possible example in the automotive industry is with the terms “Drivetrain” (T1) and “Powertrain” (T2). Drivetrain’s meaning (M1) is more precise and could be included in the meaning of Powertrain (M2).

The granular-homonymy merges the concept of hypernymy/hyponymy and homonymy. Thus, the term T1 has two meanings and the meaning M2 is the “hyponym” of M1, M1 is obviously the “hyponym” of M2. M1 has a meaning that is more precise than M2. In other words, M2 is a generalized meaning of M1. A possible example is around the term “Wheel” (T1), for a team or enterprise “wheel” can mean the “rim and tire” (M1), but for another entity, “wheel” could erroneously refer to a “rim” (M2). M1 is more precise than M2.

The partial-synonymy involves two terms and two meanings. The term T1 is associated to the meaning M1 and likewise for T2 and M2. This case introduces the concept of “inclusion” described previously as the meaning M1 and M2 share a common “space”, which is a common but partial meaning. In the automotive industry, there is the “Powertrain” (T1) that represents the engine, the gearbox and other components that produce and transmit the power to the driving wheel (M1), and there is the “Motor” (T2) which only provides power to drive the vehicle. Both T1 and T2 have a common meaning: giving torque to the driving wheel.

The next case is not so far from the previous concept. A term T1 has two meanings M1 and M2 that are not totally disjoined. An “inclusion” could be noticed between M1 and M2, thus a common but partial meaning is shared. This case is the so called partial-homonymy. This is the case for the term “Bending” (T1) which could mean to bend a stainless steel sheet (M1) or to apply a force on a beam in a material resistance test (M2). The common meaning is the realized movement.

The total-synonymy and the total-homonymy are the “conventional” form of the definitions. The total-synonymy represents two terms with the same meaning and the total-homonymy represents one term with two different meanings. Note that terms and meanings are completely separated in these cases. An example of a total-synonymy are the terms “Engine” and “Motor” that both have the same meaning: the device/system able to drive another system. For the total-homonymy in a mechatronic context could be the term “Resistance” which, depending on the domain, can mean an electrical/electroical component or the mechanical resistance of a material.

This first sub section propose six different cases and their respective examples in order to identify and solve semantic problems that could be created within teams, between teams, projects and companies. These six cases are mainly based on homonymy and synonymy but also involve hypernymy/hyponymy and the concept of “inclusion”: a shared meaning. This work will be part of the data extraction process that is described in the next sub section.

4.2. Data extraction process

The second sub-proposition is the two-step process to support the extraction from the third-parties systems via a developed connector, to process the extracted data and to store the result in a knowledge data-base. This process is represented by the lower part on Fig.3.

This connector is developed by a partner in our research project. Most current connectors between two APIs are developed by software engineers and are not understood by mechanical and electronic engineers that make these connectors unable to perform as expected. Whereas, this new connector is wanted to be intuitive and usable by non-computer scientists. This will be achieved through a development framework that will guide discipline-specific users toward the implementation of a custom-made connector suited for their discipline. Thanks to this framework the connector is understandable by non-computer scientists.

The first step of the proposed process is the extraction. This extraction through the connector will extract data from a discipline software such as CatiaV5⁵, Creo 2.0⁶, Nastran⁷, AutoCAD⁸, Solidworks⁹, Dymola⁴, etc. This connector and the extract step will not extract all the data of the files. Depending of the discipline and the type of model: CAD/DMU, 0D/1D and 2D/3D simulations, the extracted data will not be the same and some data are more relevant than others for capitalization and future projects. Thus, our extraction step will target, depending on the type of model, some specific data that could be reused in further projects. These data are identified to mainly be parameters and rules. A major part of our research work is to

![Fig. 2. Six different cases of semantic problems](image-url)
determine which data is relevant and whether the data can be transformed and reused.

Besides, metadata are unavoidable and also add a value to the process by taking into account a part of the context and could facilitate a mapping between the different data. The date of creation, the version number, an existing link with another file are examples of metadata that participate in building the context and help in creating the mapping. This idea of a mapping between the different sources of data will be detailed in a future work.

Generic data and metadata that are expected to be extracted are first described in a data model. This data model is the basement of the connector: its map to access to data and metadata. Then, the connector, depending on the discipline software, will instantiate the specific data and metadata under an instantiated model format. The output of the “data extraction” step is thus a set of instantiated models.

The processing step is the final step of our process that will try to map links between instantiated models. It is expected that most links will be created automatically thanks to an algorithm based on metadata. Although, in some cases, links between two models will need to be specified by a user. As an example, it is envisioned to map an instantiated model extracted from a CAD part with an instantiated model of a 3D simulation part if the 3D simulation part is derived/simplified from the CAD part. A temporal and hierarchical link exists between them. A second example could be a versioning mapping: if an instantiated model 1 is the nth version of instantiated model 0, the two models should be mapped.

A second step in processing data is to create set/clusters of parameters and rules that are specific-discipline-focused. These parameters are identified as interdependent and are a form of knowledge. A simple example could be the mapping between a radius and a diameter by a mathematical rule.

From a technical aspect, the data model could be formalized under a XSD model format, and the instantiated model is thus an XML model.

During the final step, the results are exchanged with the knowledge data base that will store the set/clusters of parameters and rules into the appropriate form for a further use.

The next section proposes to use the semantic dictionary as a support/tool to add a value to the mapping step.

4.3. Semantic dictionary : added value

In previous sub section, a semantic dictionary including six semantic cases and a data extraction and processing process are presented. This section present the added value of merging these two propositions. The final process is illustrated by the whole Fig. 3.

It is envisioned the semantic dictionary as a support for data processing and more precisely for data mapping. The semantic dictionary allows the reconciliation between different terms and meanings that are stored in instantiated models to create a mapping thanks to the six cases. These terms and meanings were employed during the project by teams and companies. This automatically creates links that will ease and consolidate the global mapping. The semantic dictionary is thus a support for the mapping.

Finally, the added value of the proposed process is its purpose. This process is established for a KBE approach and to complete the data base with experience of past projects from different sources from various software. The semantic dictionary bring an additional value to the mapping.

5. Use-case

To demonstrate the value of our proposition, a focus is made on the design of a two-part mechanical assembly of a rim and its hub. It is considered the design of the rim is done by entity A – an enterprise or a team, and the hub is designed by entity B – an enterprise or a team. This division that often occurs in industry, introduces interfaces and possible semantic variabilities.

To illustrate the use-case, Table 1 lists three characteristics on which a semantic variability can occur.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Rim (A)</th>
<th>Hub (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotational speed</td>
<td>Rpm</td>
<td>Rad/s</td>
</tr>
<tr>
<td>Material</td>
<td>AlSi10Mg(Fe)</td>
<td>43400</td>
</tr>
<tr>
<td>Feature 1 Name</td>
<td>Drill</td>
<td>Hole</td>
</tr>
</tbody>
</table>

If a sensor is implemented, entity A works in rpm, whereas entity B deals with Rad/s. In the data extraction, the semantic dictionary will identify the two units as two terms of a total-synonymy for the meaning “rotational speed”. Both units will be linked by a mathematical expression.

About the material, in industry different denominations exists: here it’s the chemical expression and the numerical expression for an aluminum alloy. Both are two terms of a total-synonymy of the meaning “Aluminum Alloy”. As a consequence, the two terms that are contained in two different XML models are linked by the semantic dictionary.

Finally, an interface-feature is defined by entity A as a “Drill” whereas entity B prefers to use “Hole”. Although, both name the same feature. It is the total-synonymy case and the semantic dictionary should recognize and map them together as two terms which describe feature 1.

The use-case demonstrates that the semantic-dictionary will help to map different characteristics in an industrial case study where data are processed into information that will be uploaded in the KBE agnostic solution.
6. Conclusion

This paper proposes an envisioned data-extraction process for multidisciplinary product development in a routine design and KBE context. The proposition is, for present, mostly a global process and architecture because the research project is in its early stages. This process aims to fill the gap between third-parties systems and knowledge data base. Different sources of data are envisioned in these third-parties systems. Moreover, the value of meta-data linked to the extracted data is highlighted. Semantic issues are also handled through the state-of-the-art and the Ontology, but also through the six semantic cases proposed. This semantic dictionary is a support for the data-extraction process. Our future work will explore the possibility of extracting data in non-formalized formats. These data are the support of collaboration and is often set aside. Non-formalized data could be found in mail/e-mail exchanges, meeting reports and notes of phone exchanges as example. The semantic dictionary will thus try to reconcile terms from third-parties systems and non-formalized formats and exchanges to catch the collaborative value and capitalize on past projects. This whole set aims to enrich the knowledge data-base. During the operational phase, the semantic dictionary will also be useful by facilitating to search information in the knowledge base: reconciliation of terms and meanings.

Acknowledgements

The authors are grateful to the MMe project supportive partners, and the academic and industrial partners. A special thanks is addressed to all consortium partners for their valuable contribution during the development of ideas and concepts.

References