NATURAL LANGUAGE PROCESSING AND BIM IN AECO SECTOR: A STATE OF THE ART

GIUSEPPE MARTINO DI GIUDA, MIRKO LOCATELLI, and ELENA SEGHEZZI

ABC Department, Politecnico di Milano, Milano, Italy

The research provides state-of-the-art theories, methods, and applications of Natural Language Processing in a BIM approach to define advantages, weaknesses, and potential developments. Traditionally, the design and construction process requirements are expressed through verbal qualitative evaluations instead of using numerical and structured data. Information modeling and management methods can hardly manage requirements expressed by means of qualitative and unstructured expressions. This paper aims to define the state of the art on Natural Language Processing applications for the numerical translation of basic design requirements in the AECO sector. The major advantages of this research would be the optimization and automation of numerical definition and validation of requirements in a structured data form. Structured data can, in fact, be easily managed in a BIM approach. NLP supporting information modeling methods allows the application of requirements engineering and management techniques to handle construction processes from a data-driven perspective. An investigation of the setting of a decision support system, based on NLP, for the definition and validation of requirements in the early stages of the construction process, is also provided as a potential further development.

Keywords: NLP, Computational linguistic, Requirements engineering, Requirements management, Building information modeling, Information management, Data driven process.

1 INTRODUCTION

1.1 Lack of Requirements Definition and Management in AECO Industry

AECO sector (Architecture, Engineering, Construction, and Property) is affected by poor requirement management, identification, and traceability that affects the entire construction process (Yu et al. 2005, Arayici and Ahmed 2006). Requirements are the basis of each project, as they define what stakeholders and users need and what the final product should satisfy.

Requirements are essential for planning, proper risk management, information exchange, and design modification control (Hull et al. 2005); they are fundamental for the preliminary-design phase and the design and construction process (Yu et al. 2010). Requirements that are well-expressed must be unambiguous, consistent, complete, feasible, neutral to design solutions, traceable, indispensable, synthetic, and verifiable (Young 2004).

1.2 AECO Document-based Information Flow

In recent years, the AECO industry is experiencing a radical change due to the increasingly pervasive application of information modeling. However, information exchange is based on documents, especially during the preliminary phase, and AECO information flow can still be
considered document-based and not data model-based. Semi-structured and unstructured documents are an essential part of the overall project information resources (Opitz et al. 2014). In the AECO industry, requirements are defined and exchanged in natural language through text documents; natural language is pervasive. It is probably the best and richest form of knowledge representation and, therefore, cannot be replaced as the method of requirement definition during the briefing phase. However, this richness makes it difficult to automatically manage the knowledge in narrative documents (Mich and Garigliano 2014). Engineers and architects need semi-formal and formal models to work with; on the contrary, projects start with the definition of textual requirements in natural language, which may be inadequate because they are incomplete or ambiguous. The operation of identifying and translating textual requirements into semi-formal or formal requirements is usually manual, therefore error-prone (Arellano et al. 2015).

In requirement definition and management, some research has applied Natural Language Processing methods to overcome these problems (Ryan 1993, Ambriola and Gervasi 1997, Arellano et al. 2015). The complete definition of the initial requirements is fundamental for creating information models and information management processes; NLP can be performed to translate natural human language into structured digital entities. Aiming to better explain current and future implementations of NLP for information modeling in the AECO field, an overview of NLP's basic theories and latest developments is provided.

1.3 Brief History and Development of NLP

NLP is an artificial intelligence (AI) subfield that aims at making computers able to process natural human language (Cherpas 1992, Zhang and El-Gohary 2015).

Summarizing, NLP is a multidisciplinary field that involves humanistic knowledge, mathematical-statistical, and computer science competencies, aiming to process natural human languages using computers (Lenci et al. 2005). By definition, natural human language is considered ambiguous and mutable. On the other hand, machine language is structured, formal, univocal, and based on the bit (I/0), so that computers represent linguistic objects in non-linguistic ways (Espunya i Prat 1994).

The history of computational linguistics has seen the succession of two main approaches: Rule-based approach and Corpora/Statistical-based approach. The main innovations and changes in the NLP field until the late 90s are shown in the timeline in Figure 1.

<table>
<thead>
<tr>
<th>First computer applications to literature texts 50s (Padre Buss)</th>
<th>The first electronic corpora Mid-60s (Francis &amp; Kucera)</th>
<th>Symbolic Models, Logic &amp; Artificial Intelligence, Natural Language Understanding 60s and 80s (Minsky, Schank, Winograd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generative Grammar is born 1957 (Chomsky)</td>
<td>Development of Corpus Linguistics and Statistical NLP 60s and 80s (Leech, Sinclair, Herdan)</td>
<td>Empirical NLP, Statistical NLP, Non Symbolic Models 90s (Charniak, Church)</td>
</tr>
<tr>
<td>RATIONALIST APPROACH Rule-based</td>
<td>EMPIRICAL APPROACH Corpus-based</td>
<td>MACHINE LEARNING APPROACH Rule/Corpus-based</td>
</tr>
</tbody>
</table>

Figure 1. Computational linguistic timeline, from 50s to 90s.
1.4 Recent Approach to NLP Based on Deep Learning

NLP must deal with the ambiguous and inconsistent nature of human language. The latest NLP techniques privilege the use of machine learning algorithms based on Artificial Neural Network (ANNs). To manage the ambiguity of natural language, modern NLP relies on mixed methods that involve Rule-based, Corpora-based and Machine Learning (ML) approaches based on ANNs, that allow automatic modeling through learning processes (Callison-Burch and Osborne 2004). NLP techniques based on ANNs seem to have more chance of success due to various reasons. ANNs’ characteristics are reported in Table 1.

Table 1. ANNs’ potential characteristic and applications within the NLP field.

<table>
<thead>
<tr>
<th>ANNs’ characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANNs can manage noisy and ambiguous data from the human natural language and they are very similar to the human method of learning (Rumelhart et al. 1987).</td>
</tr>
<tr>
<td>ANNs can overcome lack of data, which can be summarized in the phrase &quot;Fill in the blank&quot;.</td>
</tr>
<tr>
<td>ANNs are very powerful pattern recognizers and classifiers (Jain and Pathak 2014).</td>
</tr>
</tbody>
</table>

NLP based on ANNs can translate natural language documents into formal entities and resources (Callison-Burch and Osborne 2004). However, ANNs do not explain their structure or the connections between the neurons that are created during the training phase: this phenomenon is called the black-box effect (Waziri et al. 2017).

2 NLP AND BIM

2.1 AECO Application Fields

Table 2 shows the main studies classified by field of application and sorted chronologically. The thematic and chronological subdivision allows to investigate the presence or absence of application cases and the evolution of the research in various phases of the construction process.

2.2 NLP and BIM for Requirements Engineering and Management in AECO industry

Research and applications on Requirements Engineering (RsE) in Systems Engineering (SE) are limited and not detailed in the AECO sector; there is also a lack of researches and application cases on Requirements Management (RsM) field (Yu and Chan 2009). Different applications of Natural Language Processing to the Requirements Engineering sector have been tested (Ambriola and Gervasi 1997, Ryan 1993). Applications include the use of NLP to extract ontologies from requirements in a specific knowledge domain and the use of NLP to verify consistency and/or completion of requirements (Arellano et al. 2015).

In the AECO sector, NLP can support the client to express and define his own requirements through formal entities. The modeling process of the client's requirements could increase the capacity of managing the construction process. The application of NLP during the definition, verification, and validation of project requirements can mitigate the risk of overcoming costs, time and budget, and support quality monitoring. An automated or semi-automated translation of narrative information into formal entities is a mandatory step for the application of Building Information Modeling approaches, which can be extended to an Information and Knowledge modeling and management by the semantic modeling of the requirements through NLP.

In a truly digitalized AECO industry, semantic information must be digitized and codified to make it unequivocally machine-readable; from this point of view, current BIM methodology shows limits and issues during the preliminary phase (Simeone and Cursi 2016). Requirements
identified and defined in the briefing phase can be modeled and shared through semantic modeling methods in order to achieve a fully data-driven construction process (Beetz et al. 2005). The semantic enrichment of the information modeling process passes through applications derived from the semantic web. AECO sector sees an amount of experimental applications of semantic modeling of the knowledge domain based on ontologies (Beetz et al. 2005, Pauwels et al. 2019). The definition of ontologies describes them as formal and explicit specifications of shared conceptualizations (Gruber 1995). A requirement specification has no difference to an ontology, namely an explicit and shared conceptualization of a need and as such modeled.

Table 2. Applications of NLP in AECO sector.

<table>
<thead>
<tr>
<th>AECO field</th>
<th>NLP application</th>
<th>Main goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information management</td>
<td>Automated hierarchical document classification system (Caldas and Soibelman 2003).</td>
<td>Improve information organization and access in construction management information databases.</td>
</tr>
<tr>
<td></td>
<td>Explicit semantic analysis for product model retrieval in construction industry (Liu et al. 2017).</td>
<td>Build up a retrieval system for BIM product models, improving information extraction.</td>
</tr>
<tr>
<td></td>
<td>Construction Project Key-Phrase Network system for the representation of unstructured documents (Nedeljkovic and Kovašević 2017).</td>
<td>Improve acquisition, analysis, and reuse of relevant information in an integral form visualizing valuable project facts.</td>
</tr>
<tr>
<td></td>
<td>Natural-language-based intelligent retrieval engine for BIM object databases (Wu et al. 2019).</td>
<td>Improve BIM object query using semantic information.</td>
</tr>
<tr>
<td></td>
<td>NLP to analyze construction project contracts and correspondence (Marzouk and Enaba 2019).</td>
<td>Help project parties to figure out obligations and reduce time and effort for contract analysis.</td>
</tr>
<tr>
<td>Risk management</td>
<td>Analysis of bidding documents based on NLP systems (Lee and Yi 2017).</td>
<td>Construction projects bidding risk prediction and evaluation.</td>
</tr>
<tr>
<td>Requirement management</td>
<td>NLP to detect defects in the requirement documents (Ferrari et al. 2018).</td>
<td>Apply and test NLP system in the industrial settings of a railway signaling manufacturer.</td>
</tr>
<tr>
<td></td>
<td>Automated requirements identification from contracts through a classification model based on Naïve Bayes (Le et al. 2019).</td>
<td>Support project definition determination automatically separating requirement statements from non-requirement statements.</td>
</tr>
<tr>
<td>Automated compliance</td>
<td>Semantic machine learning-based text classification algorithm for classifying clauses and sub-clauses (Salama and El-Gohary 2016).</td>
<td>Enhance Automated Compliance Checking (ACC) in AECO industry.</td>
</tr>
<tr>
<td>checking</td>
<td>NLP and deep learning-based approach, translating building regulations to computer-readable format (Song et al. 2018).</td>
<td>Support Automated Rule Checking activity in AECO industry.</td>
</tr>
<tr>
<td>Construction management</td>
<td>Classification techniques for assigning work descriptions to task groups based on construction vocabulary (Martinez-Rojas et al. 2018).</td>
<td>Automatically classify semantic information about nature of the work tasks supporting project managers to manage and achieve project goals.</td>
</tr>
<tr>
<td></td>
<td>Text mining and NLP to analyze construction site accident (Zhang et al. 2019).</td>
<td></td>
</tr>
</tbody>
</table>

3 CONCLUSION AND FURTHER DEVELOPMENTS
The application of NLP and BIM methodology can optimize and automate the numerical definition, validation, and management of requirements in a structured data form. NLP
supporting information modeling methods allows the application of requirements engineering and management techniques to handle the preliminary phase of the construction processes from a data-driven perspective. As shown above, ontologies are formal and shared entities of a conceptualization. An NLP system could recognize and translate requirement specifications into ontologies. Structured information and knowledge, like ontologies, can be managed in a BIM process from the initial phase of the construction process. With this perspective, NLP and BIM methods can enhance semantic modeling of requirements definition and management of the domain knowledge. Modeling the semantic information contained in text documents through NLP methods during the preliminary phase could avoid or reduce errors during the operation of identifying and translating textual requirements into semi-formal or formal requirements. A decision support system based on ontologies can also be applied to manage the information flow in a data-driven way, overcoming the traditional document-based information flow of the AECO industry, enabling at the same time the application of Requirements Engineering and Management.

References


