

# EXPLORING BIM AND NLP APPLICATIONS: A SCIENTOMETRIC APPROACH

MIRKO LOCATELLI<sup>1</sup>, ELENA SEGHEZZI<sup>1</sup>, and GIUSEPPE MARTINO DI GIUDA<sup>2</sup>

<sup>1</sup>*ABC Dept., Politecnico di Milano, Milano, Italy*

<sup>2</sup>*Dept of Management, Università degli Studi di Torino, Torino, Italy*

In AECO (Architecture, Engineering, Construction and Owner-operated) industry information are mainly defined and exchanged in natural language through textual files and documents. On the contrary, a Building Information Modeling (BIM) approach requires machine processable data and information. Natural Language Processing (NLP) allows to process textual information into structured format. About BIM in AECO industry, and NLP in different sectors, several literature reviews have been conducted. However, none of them highlighted the possible connections between the two topics. This study provides a scientometric analysis aiming to investigate possible combined applications of BIM and NLP. A quantitative literature review approach is employed, using data visualization and science mapping applied on bibliometric meta-data. The performed analysis uncovered possible directions for further research on NLP and BIM combined applications in AECO sector. Most active authors, key research patterns, and institutional affiliations are identified. The main applications areas of a combined NLP and BIM approach in AECO are: Information Retrieval and Information Enrichment of BIM models, Automatic Compliance Checking, and Safety and Risk Management. The keywords pattern analysis highlighted the main tools which allows to link Semantic BIM and NLP methods and technologies, i.e. Ontology and Machine Learning algorithms. The scientometric analysis also reveals a gap related to the Preliminary design and Requirement definition phases, highlighting a possible research area not covered by the Academia as of now.

*Keywords:* Natural language processing, Building information modeling, Bibliometric analysis, AECO sector, Semantic, Artificial intelligence, Model based approach.

## 1 INTRODUCTION

### 1.1 Natural Language Processing and Semantic BIM

Natural Language Processing (NLP) is a branch of Artificial Intelligence (AI) devoted to the automatic analysis and representation of human language by formal and machine-readable language (Young *et al.* 2018).

The NLP task, i.e. development of algorithms and systems that enable computers to perform tasks involving human language, is a considerable challenge in AI field. NLP aims at solving the issue of making the human language machine readable and processable. Regarding semantic and linguistic entities also the Building Information Modeling (BIM) approach has progressively shown limitations in terms of semantic modeling and representation (Cursi *et al.* 2017). In order to overcome the semantic information-related issue, researches and studies about semantic BIM enrichment are carried out (Simeone and Cursi 2016). Semantic BIM aims to represent knowledge rather than merely data and information adding a new kind of information to the BIM

approach (Jeong 2008). From these assumptions it is clear the potentiality of NLP to manage the semantic information in a BIM approach in order to overcome the BIM semantic issue.

## 1.2 Document-Based and Model-Based Approach

In AECO (Architecture, Engineering, Construction and Owner-operated) industry information is traditionally defined and exchanged by semi- or un-structured documents (Opitz *et al.* 2014). The construction process deals with several, different and complex information that is exchanged and modified by the actors involved and much of it is captured using documents. From this perspective the AECO industry can be defined as a document-based industry. Meanwhile, the rising of the BIM methodology tends to shift the sector toward a model-based approach, which is focused on the development of models. The coexistence of a double approach, i.e. the document-based and the model-based one, leads to the separation of the informative source which can be counterproductive causing information duplications and losses (Opitz *et al.* 2014). The application of NLP system in order to process and translate semantic information into machine-manageable format in a Semantic BIM environment could open numerous applications and benefits regarding information and project management in the construction field.

About Semantic BIM in AECO industry (Godager 2018), and NLP in different sectors, some major literature reviews have been conducted (Sparck J. 1994, Khurana *et al.* 2017). However, none of them try to uncover and highlight the possible connection between BIM and NLP methods and technologies in order to manage semantic data and information. The presented bibliometric analysis tries to address this gap producing a scientometric literature review to explore the existence and dissemination of methods and applications of a BIM and NLP combined approach in AECO sector.

The work is divided into two main parts: the first one describes the method and the first analysis about the temporal trends of the two topics NLP and Semantic BIM individually. The second part investigates the two intertwined topics presenting the actual scientometric analysis.

## 2 PART 1: NLP AND SEMANTIC BIM TEMPORAL TRENDS

### 2.1 Method for Data Collection and Bibliometric Analysis

The Bibliometric analysis was conducted in September 2020 querying two online databases (DBs), i.e. Scopus and Web of Science (WoS), and followed the standard process of the systematic literature review (Moher *et al.* 2015). Specific criteria were applied to perform the analysis: only available Full-English-text articles have been considered, and logic keyword strings have been defined, also by the use of wild characters \* and Boolean operators (“AND” and “OR”) that allowed to refine the queries of the two DBs.

### 2.2 Semantic BIM and NLP Independent Queries

As regards the first part of this study, in order to uncover temporal trends of Semantic BIM and NLP topics individually, a preliminary investigation was conducted with the method described above, and the following two logic keywords strings were used to query the two DBs: I) a first string for the Semantic BIM topic (((“BIM” OR “building information model\*”) AND “semantic”); II) a second string for the NLP topic (“NLP” OR “Natural Language Processing”). Regarding the Semantic BIM analysis, the time period chosen for the data query went from 2010, when the Semantic BIM keyword first appeared in the DBs, to 2019 (excluding 2020 publications). The time period chosen for the NLP topic analysis went from 1988, since 1980s-1990s which is known as the period of statistical NLP revolution (Sparck J. 1994), to 2019. All

the subject areas were included in the query in order to define a general trends of NLP topic in the scientific community.

### 2.3 Results and Discussion

Considering the annual scientific production represented in Figure 1, an increasing interest in NLP system and technology is noticeable in recent years. The annual growth rate, considering the last period (2018-2019) is equal to 21,52% in WoS DB and 28,69% in Scopus DB.

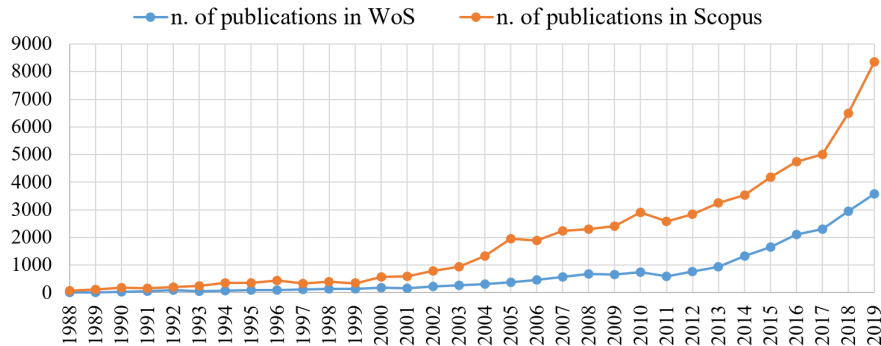


Figure 1. Scientific production 1988-2019 about NLP topic.

An analysis of the distribution of the publications is performed (Figure 2), classifying the articles according to the subject areas common to both Scopus and WoS DBs. Figure 2 shows that in both DBs publications and applications of NLP systems and technologies mainly belong to Computer Science, Engineering and Medicine domains, showing a good amount of studies in the Engineering field.

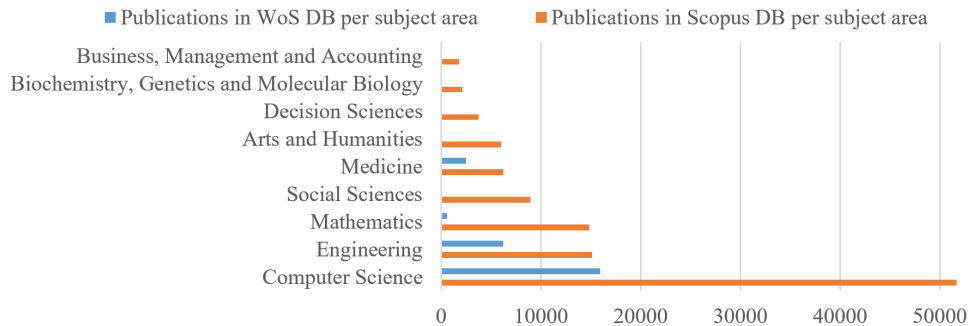


Figure 2. Publications about NLP topic divided per subject area.

From the result of the preliminary analysis it is possible to state that NLP systems and technologies seem to be a trending topic and a developing technology in several fields apart from the Computer Science specific knowledge domain. NLP systems seem to have become a key competitive element in modern information technology, with applications in theoretically all sectors and areas that deal with language and semantic data. The diffusion of NLP systems to sectors beyond the Computer Science field suggests a sufficient maturity level of the technology.

Likewise the annual scientific production about Semantic BIM, presented in Figure 3, shows an increasing interest in AECO sector related subject areas, with an annual growth rate (2018-2019) equal to 32,39% in WoS and 28,07% in Scopus DBs.

As a conclusion, both topics NLP and Semantic BIM are considered by the scientific community as valid topics of investigation and research.

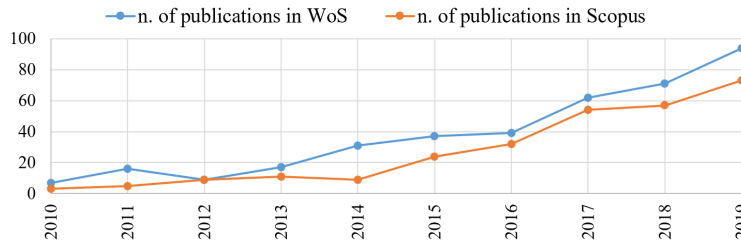


Figure 3. Scientific production 2010-2019 about Semantic BIM topic.

### 3 PART 2: BIBLIOMETRIC ANALYSIS OF COMBINED NLP AND BIM

#### 3.1 Semantic BIM and NLP Combined Application Query

Regarding the second part of the study, the same criteria described above have been applied. Moreover, an additional preliminary step was performed in order to define the keyword string for the combined application of BIM and NLP: a limited number of articles corresponding to the boundaries of the research field in BIM, NLP and the combined application of the two in AECO sector, was selected and analyzed. As a result, a set of proper keywords was extracted, as a base for the following logic keywords string, used to query both Scopus and WoS DBs: (((("NLP" OR "natural language processing")) AND (("BIM" OR "building information model\*")))).

The collected set of publications was cleaned from duplicated papers, and filtered by subject areas, excluding the knowledge fields not related to AECO industry. Then, a further screening phase was manually performed on the sample by two of the co-authors, first by title and abstract, and then by full-text reading. This enabled a further identification of the main research areas discussed in the following section. The double-check analysis allowed the improvement of the classification methodology of the sample into application fields.

The final sample of 33 publications was analyzed to perform the bibliometric analysis by means of Bibliometrix, an R package (Aria and Cuccurullo 2017), in order to visualize and uncover the possible connections between BIM and NLP methods and technologies.

#### 3.2 Results and Discussion

##### 3.2.1 Main research areas in AECO sector

The sample of 33 articles is divided into 8 main fields of application within AECO sector (Figure 4). The field of Automatic Compliance Checking has the highest number of articles. A reason can be the structure itself of regulations, since building codes documents are well structured into formats and provide a good amount of information to derive rules for extracting and translating semantics into machine-readable language, thus facilitating the application of NLP system for Automatic Compliance Checking. Furthermore, the development of NLP-based systems to convert regulatory information is an active field of research. The second and third application fields regard Information Retrieval and Information Enrichment of BIM objects and models. Other applications of NLP and BIM are in the field of Safety and Risk Management, extracting,

translating and optimizing safety issues, by predicting more likely dangerous situations during the construction operations. NLP has also been used to analyze contracts extracting risky clauses from tender documents supporting risk management activities. However, no articles seem to be related to Preliminary design or Requirement definition.

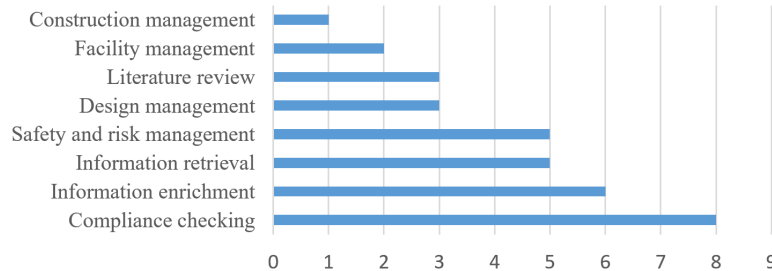


Figure 4. Main application fields of NLP and Semantic BIM.

### 3.2.2 Authors, keyword patterns and geographical provenance

In order to uncover the most active authors, key research patterns, and institutional affiliations, a data visualization task on the collected bibliometric data is performed.

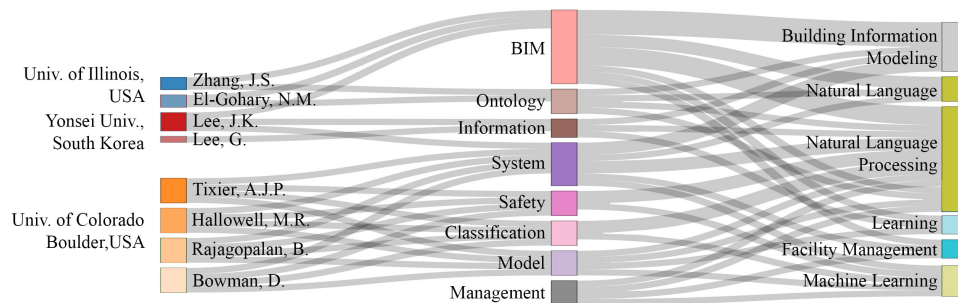


Figure 5. Three fields plot graph. Authors, Keywords plus and authors' keywords.

The three-fields plot graph in Figure 5 shows on the left the most active authors in the field, as well as their institutional affiliations. Authors are grouped by their affiliations using similar colors. Two main groups are identified on the basis of being co-authors of some publications or working on the same topics. In the middle of the graph, the Keywords plus are visualized. Keywords plus are terms that appear in the titles of an article references, but do not appear in the title of the article itself (Garfield 1990). Finally, on the right the authors' keywords are visualized. The visualization of the links between Keywords plus and authors' keywords helps identifying connections among topics and possible keywords patterns. The Keyword plus Ontology connects the BIM topic with the NLP topic. For that reason, it can be identified as a bridge term since it could represent the Information Technology tool to fill the gap between BIM and NLP fields. The Safety and Management terms can be recognized as active application fields in AECO sector implying Machine Learning techniques and NLP, but they do not involve the BIM topic. On the other hand, the Keyword plus BIM is strongly connected with the authors' keyword Natural Language which can be linked with the Semantic BIM topic and the NLP topic itself. The Keyword plus BIM is also linked with the terms Learning and Machine Learning which represent the most suitable kind of algorithm used in NLP tasks. Finally, as regards the

countries where the most active authors work, United States and South Korea seem to have the highest scientific production about the topics analyzed.

#### 4 CONCLUSION

The first part of the bibliometric study about temporal trends of NLP and Semantic BIM shows the increasing interest of the scientific community over the two individual topics in several fields beyond the Computer Science knowledge domain. The diffusion to several sectors suggests that NLP is not a niche technology anymore. The second part of the study highlights the main applications of a combined NLP and BIM approach in AECO sector, which seem to be pervasive in several fields and tasks, i.e. Information Retrieval and Information Enrichment of BIM objects and models, Automatic Compliance Checking, and Safety and Risk Management. Despite this, no articles seem to be related to the Preliminary design or Requirement definition phases, representing possible research areas not covered by the Academia as of now. The keywords pattern analysis highlighted the main tools linking Semantic BIM and NLP, i.e. Ontology and Machine Learning algorithms.

Summarizing, the performed scientometric analysis uncovered some possible directions for further research on NLP and BIM combined applications, within the context of the document-based AECO industry. A NLP system could enable the recognition and translation of textual data into ontologies, i.e. structured information and knowledge entities manageable in a BIM process. NLP and BIM methods could be paired enhancing a full model-based approach in the sector avoiding the separation of the information into several informative sources (i.e. models and text documents) which has previously been identified as a highly counterproductive process.

#### References

- Aria, M., and Cuccurullo, C., *bibliometrix: An R-tool for comprehensive science mapping analysis*, Journal of Informetrics, 11(4), 959–975, November, 2017.
- Cursi, S., Simeone, D., and Coraglia, U. M., *An ontology-based platform for BIM semantic enrichment*, SPACE SYNTAX AND ONTOLOGIES, 2, 649–656, 2017.
- Garfield, E., *KeyWords Plus: ISI's Breakthrough Retrieval Method. Part 1. Expanding Your Searching Power*, Current Contents on Diskette, 32, 3–7, August, 1990.
- Godager, B., *Critical review of the integration of bim to semantic web technology*, 3D Spatial Information Science – The Engine of Change, 42(4), 303–308, 2018.
- Jeong, Y., *Mediating semantics for multidisciplinary collaborative design*, Berkeley University of California, 2008.
- Khurana, D., Koli, A., Khatter, K., and Singh, S., *Natural Language Processing: State of The Art, Current Trends and Challenges*, arXiv, August, 2017.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., and Stewart, L., *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement*, Systematic Reviews, 1–9, January, 2015.
- Opitz, F., Windisch, R., and Scherer, R. J., *Integration of document- and model-based building information for project management support*, Procedia Engineering, 85, 403–411, 2014.
- Simeone, D., and Cursi, S., *The role of semantic enrichment in Building Information Modelling*, TEMA, December, 2016.
- Sparck J., K., *Natural Language Processing: A Historical Review*, In *Current issues in computational linguistics: in honour of Don Walker*, Zampolli, A., Calzolari, N., and Palmer, M. (eds.), 9–10, 3–16, Pisa, Dordrecht, 1994.
- Young, T., Hazarika, D., Poria, S., and Cambria, E., *Recent trends in deep learning based natural language processing*, IEEE Computational Intelligence Magazine, 13(3), 55–75, November, 2018.