ENTAC 2024
XX ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO
Maceió, Brasil, 9 a 11 de outubro de 2024

Tradução automática de normas técnicas em modelo semântico: aplicação à Norma de Desempenho

Automatic translation of technical standards into a semantic model: study applied to the Performance Standard

Douglas Lopes de Souza
Universidade Federal de Viçosa | Viçosa | Brasil | douglas@ufv.br
Regina Coeli Ruschel
Universidade Estadual de Campinas | Campinas | Brasil | ruschel@unicamp.br

Resumo

As normas técnicas são preparadas em linguagem natural (LN) exigindo uma análise manual, propensa a erros e demandando um alto nível de capacitação para a programação dos requisitos em plataformas dedicadas à avaliação de conformidade automatizada. A consulta e interpretação otimizada às normas técnicas escritas em LN é prejudicada e incompatível com a digitização em curso na Arquitetura, Engenharia e Construção. Desta forma, este artigo descreve um método de extração de informações baseado em regras para a tradução automática de normas técnicas da construção civil tendo sido aplicado a Norma de Desempenho (NBR 15.575). A solução classifica-se na área de Processamento de Linguagem Natural da Inteligência Artificial. O resultado compreende um algoritmo baseado em regras de extração de informações gerando um modelo semântico de representação dos requisitos da ND. O algoritmo é composto por diversas técnicas de Processamento de Linguagem Natural baseadas em padrões linguísticos do português. Foi gerado o modelo semântico da Parte 1 da ND expresso em linguagem formal, semiestruturada em *Extensible Markup Language*, que pode consultado de forma otimizada ou processado em aplicações relacionadas à verificação de conformidade.

Palavras-chave: Processamento de linguagem natural. Norma de Desempenho. Modelo semântico de representação. Padrões linguísticos. XML.

Abstract

Technical standards are prepared in natural language (NL), require manual analysis, are prone to errors, and require high training for programming requirements on platforms dedicated to automated conformity assessment. The optimized inquiry and interpretation of technical standards must be improved and compatible with the digitization underway in architecture, engineering, and construction. Thus, this article describes a rule-based information extraction method for the automatic translation of technical construction standards, having applied the Performance Standard (NBR 15.575). The solution is classified in the area of Natural Language Processing of Artificial Intelligence. The result is an algorithm based on information extraction rules that generates a semantic model representing the ND requirements. The algorithm is

Como citar:
SOUZA, D.; RUSCHEL, R. Tradução automática de normas técnicas em modelo semântico: aplicação à Norma de Desempenho. In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 20., 2024, Maceió. **Anais**... Maceió: ANTAC, 2024.

composed of several Natural Language Processing techniques based on Portuguese linguistic patterns. The semantic model of Part 1 of the ND was generated, expressed in a formal language, semi-structured in Extensible Markup Language, which can be consulted in an optimized way or processed in applications related to compliance verification.

Keywords: Natural Language Processing. Performance Standard. Representation Semantic Model. Linguistic patterns. XML.

INTRODUÇÃO

A análise de normas técnicas, regimentos, instruções normativas e outros regulamentos é feita de modo manual, submetida à capacidade técnica e experiência do analista em identificar as regras aplicáveis a determinado contexto de projeto.
Diversas pesquisas relatam a dificuldade que esta prática exerce sobre a cenário da construção civil [1][2][3][4][5][6]. A análise manual é dificultada por regulamentos complexos, extensos e conflitantes a outras diretrizes [6].
Ainda que hoje seja possível a utilização de um software de apoio para verificação de projetos, ainda existe a dificuldade em converter os regulamentos em regras objetivas nestes programas. Parte desta dificuldade é dada pela ambiguidade dos textos orientados à linguagem natural (LN), direcionada exclusivamente para a interpretação humana [5][7]. Além disso, outra dificuldade enfrentada pelos profissionais é dada pela impossibilidade de o documento ser convertido em regras objetivas devido à natureza subjetiva dos termos e à estrutura de redação [6][8][9][10][11].
O processamento de linguagem natural (PLN) é uma estratégia usada em diferentes pesquisas nos últimos vinte anos para tentar extrair e converter os dados de regulamentos em fomatos capazes de serem computados em programas e rotinas computacionais. Algumas destas estratégias utilizam uma base semântica de informações, estruturada para ampliar a qualidade de captura de informações [12][13], mas também se apropriam de rotinas de PLN como expressões regulares e padrões linguísticos para capturar os principais termos do regulamento [13][14].
Como resultado, é possível fazer uma tradução do texto original do regulamento em outro formato capaz de ser processado automaticamente por programas. Existem diversas pesquisas que abordaram diferentes métodos para implementar modelos de regras. Atualmente são descritos [2] os modelos utilizados para representação e avaliação de regras dos regulamentos. Ela elenca os modelos SASE (*Standards Analysis, Synthesis and Expression*) [15], modelos baseados em regras, baseados em lógica, orientados por objetivos, modelos semânticos e baseados em ontologia.
Dentre estes modelos, o semântico se diferencia pelo seu produto. Enquanto nas outras abordagens existe o interesse em conversão das diretrizes normativas em regras explícitas para serem incorporadas em plataformas de checagem (p.ex.: Solibri, Autodesk Model Checker e KBIM), o modelo semântico cria uma representação de dados estruturados ou semiestruturados explicitando o sentido dos dados. Ao representar os componentes textuais de forma estruturada, este modelo permite que sejam inseridos rótulos semânticos que classificam os termos do regulamento. Assim que os termos são rotulados, é possível criar um formato de arquivo que pode ser compreendido tanto pelos sistemas computacionais como por seres humanos. Portanto, a criação de modelos semânticos é entendida como uma abordagem de representação que trata da incorporação de novas camadas de informação nos dados, por meio do enriquecimento semântico. Os produtos destes modelos são usualmente representados por linguagens de marcação como XML, JSON e HTML.

Neste sentido, este trabalho tem como objetivo desenvolver um método de extração de informações de regulamentos voltado para a conversão de textos em formato computável, e se utiliza da Norma NBR 15.575 como exemplo de aplicação.

MÉTODO

O presente trabalho é produto de uma *Design Science Research* (DSR) e foi estruturado a partir de uma abordagem experimental do processo de extração de informações a partir de padrões linguísticos revelados pela análise do texto normativo. Na fase de desenvolvimento, foram explorados diferentes ferramentas e bibliotecas como a Spacy 3.7.2 para Python 3.10.12 visando a criação do método de extração de informações baseado em padrões linguísticos. Já na fase final do DSR, a de análise, desenvolvemos métricas quantitativas e comparamos o resultado com outros métodos usados no presente estado da arte.

RESULTADOS

Os dados de entrada do modelo foram obtidos por meio da conversão de trechos isolados da norma que representavam as características gerais dos critérios presentes no texto em linguagem natural. Após o desenvolvimento de métodos auxiliares para processamento do texto, conversão do arquivo, em formato PDF, para TXT, foi realizada a tokenização das sentenças, ou seja, separação das frases em palavras isoladas.

TOKENIZAÇÃO E ROTULAGEM COM ATRIBUTOS LINGUÍSTICOS

A tokenização ainda consegue manter os atributos linguísticos de suas funções sintáticas, gramaticais e morfológicas na sentença (sujeito, verbo, objeto direto, pronome, numeral, etc.) (Quadro 1) que são revelados na análise linguística da sentença.

O processo de tokenização e elaboração dos rótulos linguísticos foi feita pela biblioteca Spacy. Nesta etapa foi verificado que a preparação de texto, de modo convencional com a remoção de pontuações e *stopwords,* e frequentemente adotado para a língua inglesa, não pode ser adotado em regulamentos de língua portuguesa de modo direto porque elimina expressões importantes como “porta de correr” e símbolos de unidades de medida. Esta pesquisa desenvolveu análises intermediárias comparativas que indicaram que a remoção de stopwords e pontuações não obteve incremento significativo das rotulagens linguísticas.

Quadro 1 – Exemplo de atributos linguísticos de uma sentença da Norma de Desempenho.

[('instalações', 'NOUN', 'nsubj'), ('gás', 'ADJ', 'amod'), ('devem', 'VERB', 'ROOT'), ('projetadas', 'VERB', 'xcomp'), ('executadas', 'VERB', 'acl')

Fonte: Os autores.

A partir da criação de uma matriz que registrou todas os atributos linguísticos das sentenças do trecho selecionado, foi feita a análise manual das sentenças em busca de padrões linguísticos que caracterizavam os termos desejados.

O Quadro 2 exemplifica parte dos padrões linguísticos identificados para representar referências ou menções a outras normas e regulamentos dentro do texto original.

Quadro 2 - Exemplos de padrões compostos por expressões regulares

```
[{"LOWER": "normas"},{"POS": "ADJ"}],
[{"LOWER": "normas"},{"LOWER": "específicas"}],
[{"LOWER": "normas"},{"LOWER": "brasileiras"}],
[{"LOWER": "legislação"},{"LOWER": "vigente"}],
[{"LOWER": "abnt"},{"LOWER": "nbr"},{"POS":"NUM"}],
[{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}},{"POS": "PROPN"},{"POS": "NUM"}],
[{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}},{"POS": "ADP"},{"POS": "NUM"}],
[{"POS": "PROPN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}},{"POS": "ADP"},{"POS": "NUM"}],
[{"POS": "NOUN", "MORPH": {"IS_SUPERSET": ["Gender=Fem"]}},{"POS": "PROPN"},{"POS": "ADJ"}],
[{"POS": "PROPN", "DEP":"obj"},{"POS": "PROPN"}]
```

Fonte: Os autores.

Por exemplo, o padrão "[{"LOWER": "abnt"},{"LOWER": "nbr"},{"POS":"NUM"}]", busca e classifica qualquer conjunto de palavras que possuam os atributos "nome = abnt + nome = nbr + numeral" para reconhecer menções à outras normas como "ABNT NBR 13523".

ROTULAGEM SEMÂNTICA

Ao reconhecer os elementos textuais a partir da combinação de padrões linguísticos, aplicamos rótulos semânticos pré-determiandos que classificam o elemento a ele vinculado. Por exemplo, no caso acima, "ABNT NBR 13523" recebe o rótulo "REFERENCIA", indicando que aquele conjunto de palavras e numerais deve ser reconhecido como uma referência à outro regulamento.

O Quadro 3 apresenta exemplos de rótulos atribuídos a outros termos da Norma.

Quadro 3 - Trecho do esultado da extração do Capítulo 8 da Norma de Desempenho

Frase	Objeto Semântico	Texto	Rótulo
Frase 1	Objeto Semântico 1	ABNT NBR 5419	REFERENCIA
Frase 1	Objeto Semântico 2	Normas Brasileiras	REFERENCIA
Frase 1	Objeto Semântico 3	legislação vigente	REFERENCIA
Frase 1	Objeto Semântico 4	proteção contra descargas atmosféricas	COMPONENTE
Frase 1	Objeto Semântico 5	edifícios multifamiliares	COMPONENTE
Frase 1	Objeto Semântico 6	devem ser providos	EXIGENCIA
Frase 2	Objeto Semântico 1	ABNT NBR 5410	REFERENCIA
Frase 2	Objeto Semântico 2	Normas Brasileiras	REFERENCIA
Frase 2	Objeto Semântico 3	instalações elétricas	COMPONENTE
Frase 2	Objeto Semântico 4	edificações habitacionais	COMPONENTE

Fonte: Os autores.

Assim que os rótulos foram atribuídos aos termos encontrados pelos padrões linguísticos, foram feitas análises dos resultados para calibração e readequação da composição dos padrões que representam os termos semânticos. Ao final desta etapa foi feita a avaliação dos resultados de classificação dos termos segundo uma matriz de confusão que registrou erros e acertos da rotulagem semânica (falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos) (Quadro 4).

Quadro 4 – Exemplo de avaliação de uma sentença extraída e classificada pelo método

As instalações elétricas das edificações habitacionais devem ser projetadas de acordo com a ABNT NBR 5410 e Normas Brasileiras aplicáveis.	VP	5
	FN	0
	FP	0
	VN	1

Fonte: Os autores.

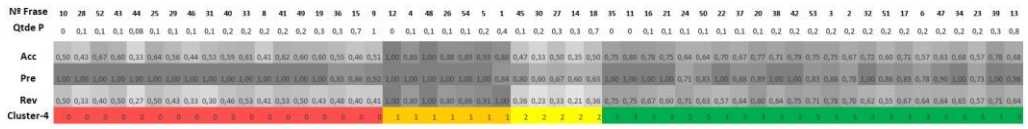
O resultado desta avaliação atingiu acurácia de 85%, precisão de 95% e Valor F1 79% para classificação de todos os termos existentes no segmento da norma usado para composição dos padrões.

Em seguida, o método foi aplicado em outro capítulo da mesma norma, ou seja, com o mesmo contexto terminológico, estrutural e de domínio do conhecimento. A aplicação do método obteve os resultados de acurácia 61%, precisão 90% e valor F1 68%, indicando prejuísos no desempenho da aplicação do modelo em outro texto em que os padrões não foram observados ou coletados previamente.

Com o intuito de identificar o motivo da perda de desempenho do modelo, foi feita um agrupamento dos resultados usando um modelo K-Means, apoiado pelo método da silhueta, para identificação da quantidade de grupos iniciais.

O modelo apontou quatro grupos distintos de sentenças separados pelos seus desempenhos nas métricas de acurácia, precisão e revocação (Figura 1).

Figura 1 - Exemplo de padrões linguísticos e seus respectivos rótulos semânticos



Fonte: Os autores.

Ao analisar as sentenças de pior desempenho (cor amarela), foi possível identificar alguns dos motivos que levaram ao baixo reconhecimento os objetos semânticos.

Um dos problemas encontrados nesta análise estava relacionado à estrutura de algumas sentenças presentes no segundo segmento de texto e ao surgimento de novos padrões, inexistentes no modelo anterior. Um exemplo de padrão que diminuiu a acurácia do sistema são frases com termos encadeados, como " instalações de de água, esgotos ou águas pluviais e estrutura, pisos e paredes".

A fim de verificar a capacidade de aplicação do modelo baseado em padrões, foi feita a comparação do modelo com Grandes Modelos de Linguagem (*Large Language Models*, LLM), considerados o estado da arte em Processamento de Linguagem Natural.

No momento da pesquisa foi desenvolvido um prompt de comando solicitando que o modelo de linguagem ChatGPT 3.5, desenvolvido pela OpenAI, encontrasse os objetos semânticos em um trecho do texto. O mesmo trecho foi analisado pelo modelo baseado em padrões linguísticos, desenvolvido por esta pesquisa. O resultado comparativo dos dois métodos frente à um padrão ouro, é apresentado no Quadro 5:

Quadro 5 - Comparação das abordagens baseadas em padrões e LLM

Gold Standard: Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).
Baseado em padrões: Devem ser previstos nos projetos a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos elementos da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).
GPT 3.5 via API: Devem ser previstos nos **projetos** a prevenção de infiltração da água de chuva e da umidade do solo nas habitações, por meio dos detalhes indicados a seguir: a) condições de implantação dos conjuntos habitacionais, de forma a drenar adequadamente a água de chuva incidente em ruas internas, lotes vizinhos ou mesmo no entorno próximo ao conjunto; b) sistemas que impossibilitem a penetração de líquidos ou umidades de porões e subsolos, jardins contíguos às fachadas e quaisquer paredes em contato com o solo, ou pelo direcionamento das águas, sem prejuízo da utilização do ambiente e dos sistemas correlatos e sem comprometer a segurança estrutural. No caso de haver sistemas de impermeabilização, estes devem seguir a ABNT NBR 9575; c) sistemas que impossibilitem a penetração de líquidos ou umidades em fundações e pisos em contato com o solo; d) ligação entre os diversos **elementos** da construção (como paredes e estrutura, telhado e paredes, corpo principal e pisos ou calçadas laterais).

Fonte: Os autores.

O resultado apresentado no Quadro 5 foi obtido pela aplicação de um *prompt* de comando por meio na interface pública e API disponibilizada pela OpenAI para acesso ao mesmo modelo.

Como ficou evidente no Quadro 5, o modelo baseado em regras e codificação explícita superou os resultados obtidos pelo LLM, sem trenamento prévio, em 53%, encontrando mais padrões com a rotulagem correta do que o modelo de linguagem.

Uma vez que alguns termos das frases da Parte 1 da Norma de Desempenho foram vinculados à rótulos semânticos da construção civil, estas sentenças foram incoporadas à uma representação estruturada de um arquivo XML (Figura 2).

Figura 2 - Exemplos bem-sucedidos da notação XML gerada pelo algoritmo

```
<FRASE>
  Os materiais de revestimento, acabamento e isolamento termoacústico empregados na face interna
  dos sistemas ou elementos que compõem a edificação devem ter as características de propagação
  de chamas controladas, de forma a atender aos requisitos estabelecidos nas ABNT NBR 15575-3 e
  a ABNT NBR 15575-5 e ABNT NBR 9442.
  <TERMOS>
    <REFERENCIA>ABNT NBR 15575-3</REFERENCIA>
    <REFERENCIA>ABNT NBR 15575-5</REFERENCIA>
    <REFERENCIA>ABNT NBR 9442</REFERENCIA>
    <EXIGENCIA>devem ter</EXIGENCIA>
    <COMPONENTE>características de propagação de chamas controladas,</COMPONENTE>
    <COMPONENTE>materiais de revestimento</COMPONENTE>
    <COMPONENTE>isolamento termoacústico</COMPONENTE>
    <COMPONENTE>face interna</COMPONENTE>
  </TERMOS>
```

Fonte: Os autores.

A adoção do arquivo XML foi devida às análises anteriores desenvolvidas pela pesquisa que indicou a permeabilidade deste tipo de arquivo em programas de verificação automática de regras, e sua capacidade de expansão e incorporação de novos atributos.

CONCLUSÃO

O texto de um regulamento é redigido para leitura de seres humanos, e em linguagem natural. Desta forma, a capacidade de interpretação e incorporação de seus termos e conceitos em outras ferramentas é dificultada porque é completamente dependente da capaciadade humana em interpretá-los.

Nesta pesquisa foi adotado um método que conseguiu extrair os principais termos de uma parte da Norma de Desempenho e converteu as sentenças extraídas em um arquivo XML.

Apesar de não atingir altos níveis de acurácia e precisão no teste cego que aplicou o método em outro conjunto de frases da Norma, o modelo desenvolvido, baseado em codificação explícita e padrões linguísticos, obteve resultado de classificação superior ao modelo ChatGPT3.5. O modelo utilizado, baseado em padrões linguísticos, possui capacidade de incluir novos padrões na sua lista de busca de modo a ampliar a acurácia do modelo à medida que recebe novas sentenças e encontra mais elementos capturados anteriormente.

A pesquisa também compreende que o modelo de linguagem sem treinamento prévio em uma base de dados relacionada com o contexto da Norma, não é capaz de atingir os mesmos resultados de uma programação explícita. Neste sentido, considerando que o treinamento demanda dados rotulados em grande quantidade, inexistentes no cenário atual, o modelo baseado em padrões ainda possui capacidade de transformar os dados em um formato de arquivo com potencial de incorporação em aplicações e plataformas do setor da construção.

Neste sentido, a transformação das sentenças da norma em um arquivo XML, enriquecido com rótulos semânticos relacionados à construção civil, permitirá que estas sentenças sejam computadas, analisadas e formatadas em diferentes aplicações. A digitalização das regras existentes permitirá que a Associação Brasileira de Normas Técnicas ofereça um banco de dados comum, com regras uniformes e em um formato enriquecido e interoperável às ferramentas de avaliação de conformidade.

REFERÊNCIAS

[1] DIMYADI, Johannes; AMOR, Robert. Automated building code compliance checking – where is it at? In: CIB WORLD BUILDING CONGRESS, 19., 2013, Brisbane, Austrália. **Proceedings**… Brisbane, Austrália: Construction and Society, 2013. p. 172-185.

[2] MACIT, Sibel. Computer representation of building codes for automated compliance checking. 2014. Tese de Doutorado. **Izmir Institute of Technology**, Turquia.

[3] HJELSETH, Eilif. BIM-based model checking (BMC). In: ISSA, Raja R. A.; OLBINA, Svetlana. **Building information modeling**: applications and practices. Reston, VA: ASCE, 2015. p. 33-61. https://doi.org/10.1061/9780784413982.ch02

[4] ZHANG, Jiansong; EL-GOHARY, Nora M. Semantic-based logic representation and reasoning for automated regulatory compliance checking. **Journal of Computing in Civil Engineering**, v. 31, n. 1, 2017b. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000583

[5] SOLIMAN-JUNIOR, João; FORMOSO, Carlos T.; TZORTZOPOULOS, Patrícia. A semantic-based framework for automated rule checking in healthcare construction projects.

Canadian Journal of Civil Engineering, v. 47, n. 2, p. 202-214, 2019. https://doi.org/10.1139/cjce-2018-0460

[6] SOBHKHIZ, Soroush *et al*. Framing and evaluating the best practices of IFC-based automated rule checking: a case study. **Buildings**, v. 11, n. 10, 456, 2021. https://doi.org/10.3390/buildings11100456

[7] LI, B. *et al*. Defeasible reasoning for automated building code compliance checking. **ECPPM 2021–eWork and eBusiness in Architecture, Engineering and Construction**, p. 229-236, 2021. https://doi.org/10.1201/9781003191476

[8] KEHL, Caroline; ISATTO, Eduardo Luís. Barreiras e oportunidades para a verificação automática de regras da produção na fase de projeto com uso da tecnologia BIM. In: ENCONTRO DE TECNOLOGIA DE INFORMAÇÃO E COMUNICAÇÃO NA CONSTRUÇÃO, 7., 2015, Recife. **Anais...** Recife: UFPE, 2015.

[9] MAINARDI NETO, Antônio Ivo de B.; SANTOS, Eduardo Toledo. Verificação de regras em modelos BIM: um estudo de caso sobre projeto de arquitetura de estações metroviárias. In: ENCONTRO BRASILEIRO DE TECNOLOGIA DE INFORMAÇÃO E COMUNICAÇÃO NA CONSTRUÇÃO, 7., 2015, Recife. **Anais...** Recife: UFPE, 2015. https://doi.org/10.5151/engpro-tic2015-068

[10] KATER, Marcel; RUSCHEL, Regina Coeli. O potencial da verificação automatizada baseada em regras para as medidas de segurança contra incêndio em BIM. **Ambiente Construído**, v. 20, n. 4, p. 423-444, 2020. https://doi.org/10.1590/s1678-86212020000400481

[11] FUCHS, Stefan; AMOR, Robert. Natural language processing for building code interpretation: a systematic literature review. In: INTERNATIONAL CONFERENCE OF CIB W78, 38., 2021, Luxembourg. **Proceedings…** ITC Digital Library, 2021. p. 294-303.

[12] ZHANG, Jiansong; EL-GOHARY, Nora M. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking**. Journal of Computing in Civil Engineering**, v. 30, n. 2, p. 04015014, 2016. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000346

[13] XU, Xin; CAI, Hubo. Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure. **Advanced Engineering Informatics**, v. 48, 101288, 2021. https://doi.org/10.1016/j.aei.2021.101288

[14] ZHOU, Yu-Cheng *et al*. Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. **Computers in Industry**, v. 142, p. 103746, 2022.

[15] FENVES, Steven *et al*. **Introduction to SASE**: standards analysis, synthesis, and expression. Washington: Editorial NBSIR, 1987. (NBSIR, 87-3513).