

ABORDAGEM BASEADA EM LLM PARA A CRIAÇÃO E VISUALIZAÇÃO DE GRAFOS DE REGULAMENTOS EM DOCUMENTOS DE SEGURANÇA CONTRA INCÊNDIO

Tomás Jorge ⁽¹⁾, Jorge Magalhães ⁽¹⁾, Jéssica Reis ⁽²⁾, Rui Gavina ⁽²⁾, Pedro Oliveira ^(1,2), José Rocha ⁽³⁾, Daniel Rocha ⁽³⁾, Ricardo Santos ⁽¹⁾⁽⁴⁾, Diogo Ribeiro ⁽¹⁾⁽⁴⁾

(1) iBuilt, Instituto Superior de Engenharia do Porto, Politécnico do Porto, Porto, Portugal

(2) VN2R-Engineering Innovation Consulting LDA, Porto, Portugal

(3) Exactusensu - Consultores Associados, Lda, Porto, Portugal

(4) CONSTRUCT-iRail, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

Resumo

A aplicação eficiente dos regulamentos de segurança contra incêndio nas fases iniciais de um projeto é frequentemente dificultada pela sua linguagem não estruturada e pela dispersão por múltiplos documentos legais. Esta fragmentação coloca desafios à interpretação automática e compromete a digitalização dos processos de verificação de conformidade. Neste contexto, o presente artigo introduz uma metodologia inovadora que integra grandes modelos de linguagem (LLMs), especificamente o Gemini Flash 2.5. O principal objetivo consiste em transformar automaticamente regulamentos redigidos em linguagem natural num formato legível por máquina, representando os requisitos regulamentares sob a forma de grafos de conhecimento, com base numa ontologia de regulamentos previamente definida. A metodologia proposta foi validada utilizando regulamentos portugueses de segurança contra incêndio, demonstrando uma forte adaptabilidade a diferentes documentos regulamentares. Os resultados evidenciam uma redução significativa do esforço manual e do erro humano tradicionalmente associados às tarefas de conformidade, incorporando simultaneamente uma estratégia *Human-in-the-Loop* para assegurar a validação por especialistas em etapas críticas do processo. De um modo geral, este trabalho constitui um contributo relevante para a automatização inteligente de regulamentos e fornece uma base sólida para a verificação automática da conformidade.

1. Introdução

A segurança contra incêndio é uma questão crítica no setor da construção, com impacto direto na proteção de vidas humanas, do património edificado e do ambiente. Apesar da sua importância, os requisitos de segurança contra incêndio são frequentemente considerados apenas em fases avançadas do processo de conceção, resultando numa abordagem reativa que compromete a eficiência técnica e económica dos projetos e pode conduzir a alterações dispendiosas ou falhas críticas em situações de emergência [1]. Um dos principais entraves à integração precoce destes requisitos reside na complexidade e fragmentação da regulamentação existente. Os regulamentos, maioritariamente redigidos em linguagem natural, encontram-se dispersos por múltiplos diplomas legais e normas técnicas, originando ambiguidades

terminológicas e inconsistências semânticas que dificultam a sua interpretação automática [2]. Adicionalmente, a escassez de formatos digitais estruturados limita a aplicação sistemática de tecnologias de apoio à verificação da conformidade. Neste contexto, os avanços em inteligência artificial (IA) e processamento de linguagem natural (PLN) têm impulsionado o desenvolvimento de metodologias para converter textos regulamentares em formatos legíveis por máquina.

Diversos estudos têm abordado esta problemática, recorrendo a estratégias manuais, como a metodologia RASE [3], ou a técnicas baseadas em PLN, LLMs e ontologias para estruturar regulamentos e ligá-los a modelos BIM ou a linguagens formais de validação [2, 4-7]. No entanto, estas abordagens apresentam limitações, seja por dependerem excessivamente de processos manuais, reduzindo a escalabilidade, seja por serem demasiado automatizadas, carecendo de validação especializada.

Para ultrapassar estas limitações, este artigo propõe uma metodologia para a geração automática de Grafos de Regulamentos aplicados à segurança contra incêndio, no contexto português, com vista à verificação automática da conformidade. A metodologia combina (i) grandes modelos de linguagem (LLMs), concretamente o Gemini Flash 2.5, e (ii) uma estratégia *Human-in-the-Loop* para validação especializada. O contributo do presente trabalho centra-se na conversão automatizada de regulamentos em grafos semânticos legíveis por máquina, na integração eficaz de automação e validação humana e na definição de uma abordagem flexível, adaptável a diferentes regulamentos e contextos nacionais.

2. Metodologia proposta

A metodologia proposta, implementada em Python, tem como objetivo converter documentos regulamentares num grafo de regulamentos estruturado, de modo a suportar processos de análise, integração e verificação automática da conformidade. Conforme ilustrado na Figura 1, a metodologia é composta por três etapas principais: conversão para texto legível por máquina, enriquecimento e conversão para um grafo RDF, e visualização.

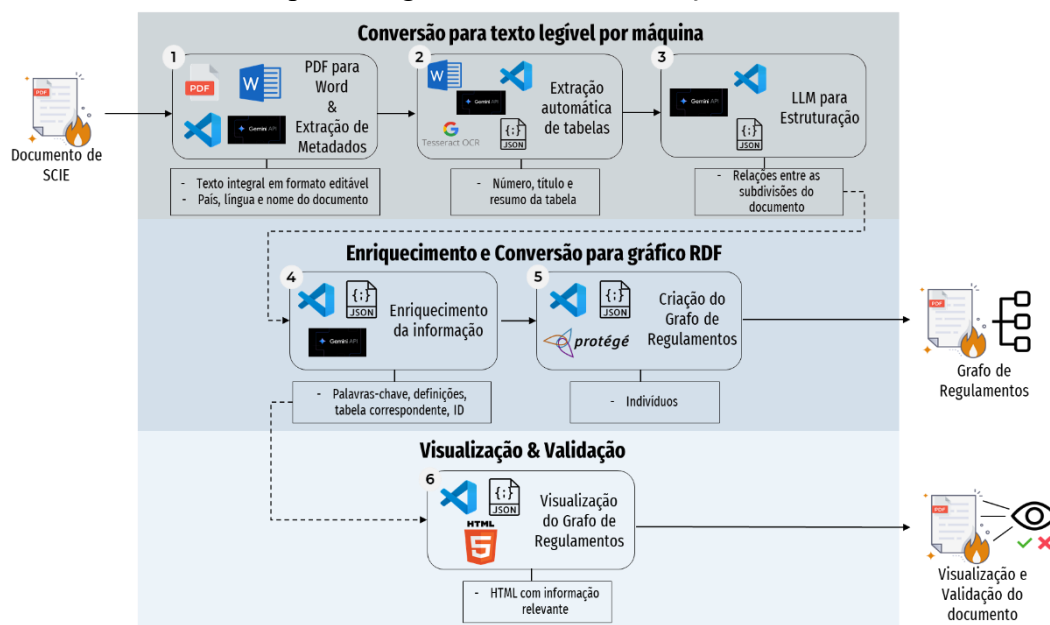


Figura 1: Fluxograma da metodologia proposta

Na Etapa 1, os documentos em formato PDF são convertidos para Word, permitindo uma extração estruturada do conteúdo e a identificação automática de metadados essenciais, como país, língua e título da legislação, através de funções em Python baseadas no modelo Gemini Flash 2.5. Estes metadados são utilizados para contextualizar corretamente as etapas seguintes. Na Etapa 2, a informação contida nas tabelas é extraída recorrendo a OCR, através do Tesseract², sendo armazenada num ficheiro JSON com a numeração, título e conteúdo. Adicionalmente, o Gemini Flash 2.5 gera um resumo interpretativo de cada tabela, permitindo capturar o seu significado semântico e apoiar a construção do Grafo de Regulamentos.

Na Etapa 3, o modelo Gemini Flash 2.5 é aplicado à classificação estrutural do documento, com base em exemplos previamente definidos. Antes de avançar, os resultados são validados através de uma interface dedicada, sendo posteriormente gerado um ficheiro JSON que representa a estrutura hierárquica do documento.

Concluída a fase de conversão, inicia-se o enriquecimento semântico. Na Etapa 4, são extraídas palavras-chave, geradas definições e estabelecidas ligações a tabelas e identificadores únicos (ID). Na Etapa 5, esta informação é utilizada para gerar grafos de regulamentos em RDF, com recurso à biblioteca rdflib e a uma ontologia previamente definida, originando uma base de conhecimento estruturada e interoperável. Os grafos RDF são visualizados no Protégé para validação *Human-in-the-Loop*.

Por fim, na Etapa 6, o grafo RDF é convertido numa interface de visualização em HTML, permitindo a exploração estruturada do conteúdo e constituindo uma base para futuros processos de verificação automática da conformidade. Para garantir a qualidade dos resultados, a validação *Human-in-the-Loop* é integrada nas Etapas 1, 2, 3 e 5.

3. Resultados

3.1. PDF para Word e Extração de Metadados

Um número significativo de regulamentos técnicos em Portugal, em particular os relacionados com a segurança contra incêndio, é publicado em formato PDF, o que dificulta a extração automática de informação apesar da boa preservação da disposição visual. Para ultrapassar estas limitações, a primeira etapa da metodologia consiste na conversão do PDF para um documento Microsoft Word, seguida de um refinamento do formato através de *Visual Basic for Applications* (VBA), resultando num ficheiro mais consistente e estruturado para processamento posterior.

A escolha do formato Word deve-se à sua maior capacidade de preservar a estrutura visual e lógica do documento, especialmente em elementos complexos como tabelas, quando comparado com formatos mais simples como *Markdown* ou TXT. Após a conversão, o documento é analisado por funções em Python baseadas em LLMs, concretamente o modelo Gemini Flash 2.5, para identificar metadados relevantes, como a língua, o país de origem e o regulamento aplicável, cuja seleção é confirmada pelo utilizador. Estes metadados são fundamentais para contextualizar corretamente o processamento subsequente e garantir uma extração de informação mais precisa. As Figuras 2a) e 2b) apresentam exemplos dos metadados obtidos.

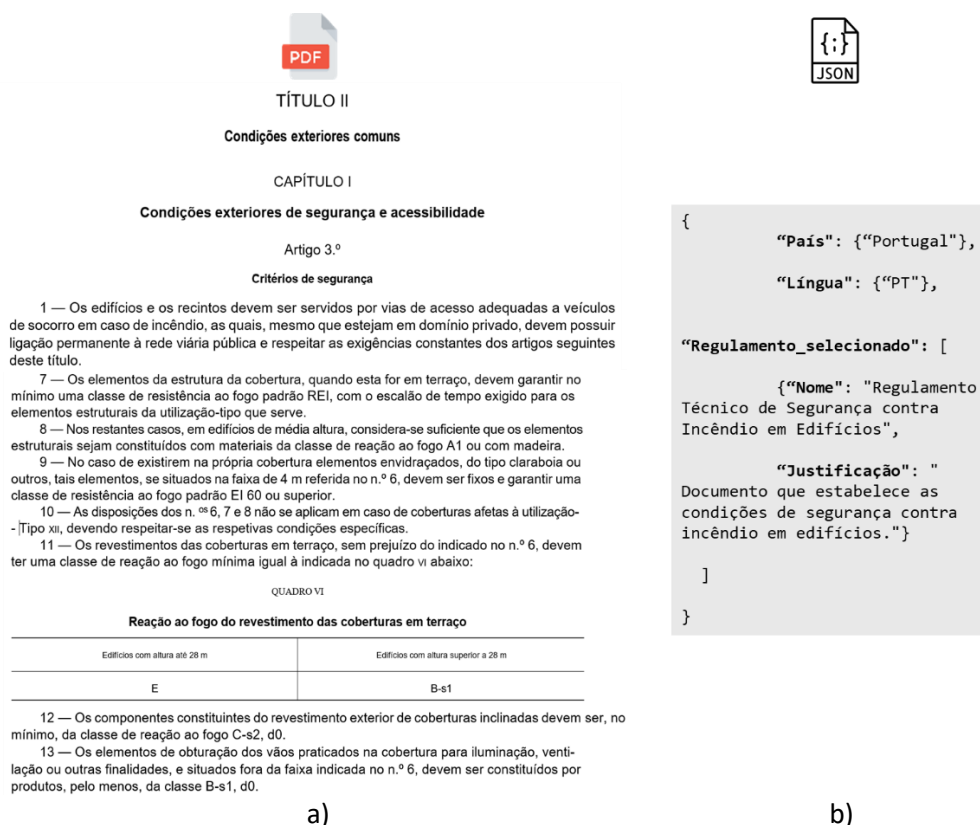


Figura 2: Exemplo do processamento do documento na primeira etapa: a) regulamento original em formato PDF; b) metadados extraídos através de funções em Python baseadas em LLMs.

3.2. Extração automática de tabelas

A extração automática de tabelas revelou-se uma das etapas mais críticas de todo o processo, devido à diversidade e complexidade estrutural dos regulamentos portugueses. Estes documentos incluem frequentemente tabelas com múltiplas colunas, cabeçalhos assimétricos, células fundidas e, por vezes, continuação ao longo de várias páginas, o que dificulta a sua deteção e interpretação automáticas. Para dar resposta a estes desafios, foram utilizadas técnicas que combinam OCR e processamento de texto. Atendendo ao padrão relativamente consistente das tabelas, iniciando-se com o número da tabela, seguido do respetivo título e conteúdo, foi possível implementar um processo de extração automatizado sem recorrer a métodos mais complexos. O procedimento inicia-se com a conversão do PDF em imagens de alta resolução, sobre as quais são aplicados operadores morfológicos para identificar estruturas tabulares com base em linhas horizontais e verticais. As regiões correspondentes às tabelas são depois recortadas e processadas através do Tesseract para reconhecimento de texto. Em paralelo, o ficheiro Word é analisado de modo a identificar os títulos das tabelas, que são corretamente associados às regiões extraídas. O resultado consiste num conjunto de tabelas armazenadas num ficheiro JSON que contém a numeração, o título, o conteúdo e uma indicação sobre a necessidade de revisão manual, particularmente relevante em situações de lacunas ou ambiguidades estruturais. Adicionalmente, as funções em Python baseadas em LLMs geram automaticamente um resumo descritivo para cada tabela, com base no respetivo conteúdo. Desta forma, a informação tabular é extraída de modo eficiente e estruturado, permitindo a sua posterior associação à hierarquia lógica do documento regulamentar. A Figura 3 ilustra o

processo de extração da informação tabular apresentada na Figura 2 e a sua conversão para um ficheiro JSON.

Interface de Edição de Tabelas

Total de tabelas agrupadas: 1 (de 1 extrações)

Selecionar tabela para editar:
Reação ao fogo do revestimento das coberturas em terraço (2 linhas x 2 colunas)

Reação ao fogo do revestimento das coberturas em terraço

Editor de Tabela

0	Selecionar	0	1
0	☐	Edifícios com altura até 28 m	Edifícios com altura superior a 28 m
1	☐	E	B-s1

Imagem de Referência

Re-extrair Imagens

10 -- As disposições dos n.ºs 6, 7 e 8 não se aplicam em caso de coberturas afetadas à utilização tipo x, devendo respeitar-se as respetivas condições específicas.
11 -- Os revestimentos das coberturas em terraço, sem prejuízo do indicado no n.º 6, devem ter uma classe de reação ao fogo mínima igual à indicada no quadro «a» abaixo:

Reação ao fogo do revestimento das coberturas em terraço

Edifícios com altura até 28 m	Edifícios com altura superior a 28 m
E	B-s1

Área da Tabela - Página 2

Dimensões: 1379x445 pixels

IMPORTANTE - Como Editar Corretamente:

1. Clique duplo na célula para editar
2. Apague o conteúdo antigo
3. Cole o novo conteúdo (Ctrl+V)
4. Pressione Enter para confirmar
5. Se não salvar: Clique no botão **Salvar Agora** abaixo

Dica: O salvamento é automático, mas use o botão se houver problemas

+ Adicionar Linha - Remover Linhas Resetar

Ferramentas de Edição

Salvar Agora Copiar Tabela Colar Dados Forçar Atualização

QUADRO VI - Reação ao fogo do revestimento das coberturas em terraço

Resumo (PT):

O Quadro VI especifica a reação ao fogo exigida para o revestimento das coberturas em terraço, diferenciando os requisitos consoante a altura do edifício. Para edifícios com altura até '28 m', é exigida uma classe de reação ao fogo 'E'. Já para edifícios com altura superior a '28 m', a exigência aumenta para a classe 'B-s1'.

Summary (EN):

Table VI specifies the required reaction to fire for the covering of terrace roofs, differentiating the requirements according to the height of the building. For buildings with a height up to '28 m', a reaction to fire class 'E' is required. For buildings with a height greater than '28 m', the requirement increases to class 'B-s1'.

Figura 3: Exemplo do processo automático de extração de tabelas.

3.3. LLM para Estruturação

A extração automática de tabelas revelou-se uma das etapas mais críticas de todo o processo, devido à diversidade e complexidade estrutural dos regulamentos portugueses. Estes documentos incluem frequentemente tabelas com múltiplas colunas, cabeçalhos assimétricos, células fundidas e, por vezes, continuação ao longo de várias páginas, o que dificulta a sua deteção e interpretação automáticas. Para dar resposta a estes desafios, foram utilizadas técnicas que combinam OCR e processamento de texto. Atendendo ao padrão relativamente consistente das tabelas, iniciando-se com o número da tabela, seguido do respetivo título e conteúdo, foi possível implementar um processo de extração automatizado sem recorrer a métodos mais complexos. O procedimento inicia-se com a conversão do PDF em imagens de alta resolução, sobre as quais são aplicados operadores morfológicos para identificar estruturas tabulares com base em linhas horizontais e verticais. As regiões correspondentes às tabelas são depois recortadas e processadas através do Tesseract para reconhecimento de texto. Em paralelo, o ficheiro Word é analisado de modo a identificar os títulos das tabelas, que são corretamente associados às regiões extraídas. O resultado consiste num conjunto de tabelas armazenadas num ficheiro JSON que contém a numeração, o título, o conteúdo e uma indicação sobre a necessidade de revisão manual, particularmente relevante em situações de lacunas ou ambiguidades estruturais. Adicionalmente, as funções em Python baseadas em LLMs geram automaticamente um resumo descritivo para cada tabela, com base no respetivo conteúdo. Desta forma, a informação tabular é extraída de modo eficiente e estruturado, permitindo a sua posterior associação à hierarquia lógica do documento regulamentar. A Figura 3 ilustra o

processo de extração da informação tabular apresentada na Figura 2 e a sua conversão para um ficheiro JSON.

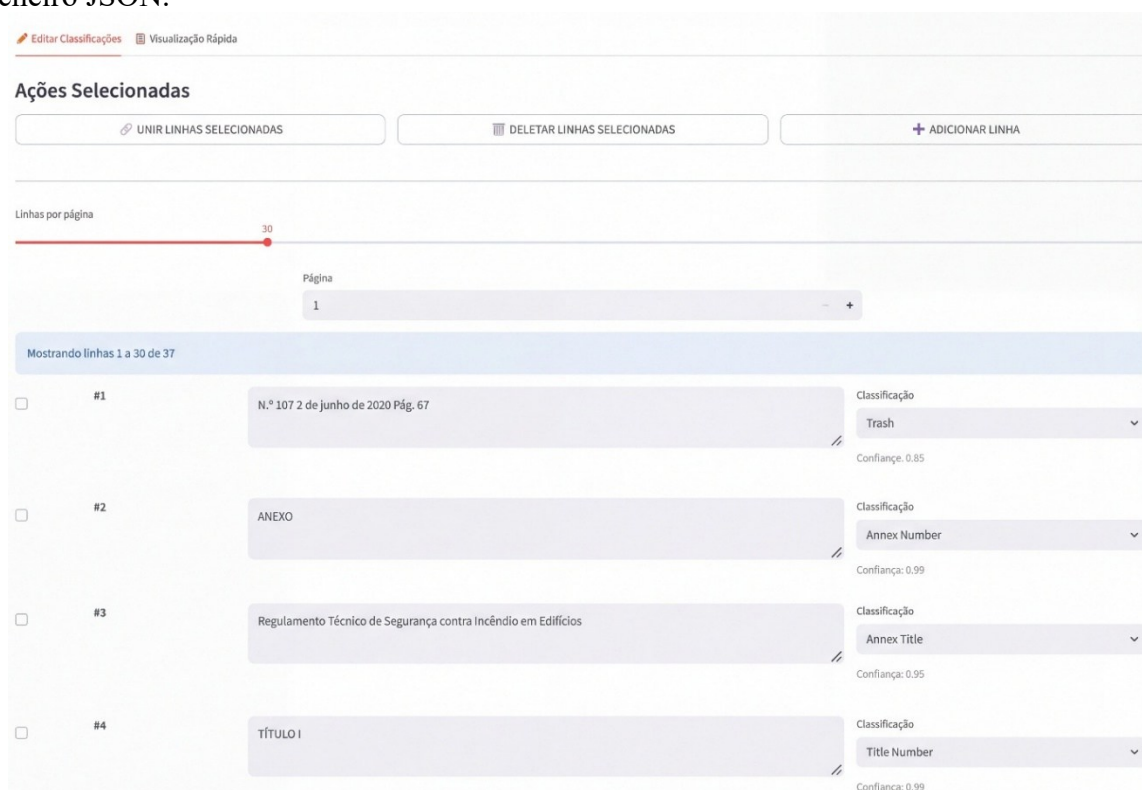


Figura 4: Interface para validação manual e ajuste das classificações automáticas.

3.4. Enriquecimento da informação

Após a classificação estrutural e a organização hierárquica do documento, inicia-se a fase de enriquecimento semântico. Esta etapa tem como objetivo complementar o conteúdo com informação adicional relevante. O processo é realizado através de funções em Python baseadas em LLMs, que analisam cada elemento no seu contexto específico.

Nesta fase, o enriquecimento incide exclusivamente sobre os elementos de nível inferior do documento, tais como artigos, parágrafos ou itens. Para estes elementos, são automaticamente extraídas palavras-chave, tanto na língua original do documento como em inglês, representando os principais conceitos abordados. Adicionalmente, o modelo gera definições concisas e contextualizadas para cada palavra-chave, com base nos metadados recolhidos na etapa inicial (por exemplo, país, língua e tipo de regulamento). Cada elemento recebe igualmente um identificador único (ID), por exemplo, “PT-Ane0_CapII_Art17_Ite2_Suba”, que indica de forma inequívoca a sua localização exata na estrutura do documento (isto é, Subitem ‘a’ do Item 2 do Artigo 17, localizado no Capítulo II do Anexo 0 de um documento português). Para além disso, os elementos que incluem tabelas recebem um indicador específico (ID_T), que assinala a presença de informação tabular associada a esse elemento.

Como resultado final, é gerado um ficheiro JSON enriquecido, no qual cada elemento textual se encontra associado às respetivas palavras-chave, definições, identificadores únicos e ligações às tabelas previamente identificadas. Este ficheiro constitui a base estruturada para a geração dos grafos de regulamentos.

```
{
  "tipo": "annex",
  "numero": "Annex 0",
  "titulo": "Regulamento Técnico de Segurança contra Incêndio em Edifícios",
  "titulos": [
    (...)
    {
      "tipo": "chapter",
      "numero": "Capítulo I",
      "titulo": "Compartimentação geral contra incêndios",
      "secoes": [],
      "artigos": [
        {
          "tipo": "article",
          "numero": "Artigo 3º",
          "titulo": "Critérios de segurança",
          "alineas": [
            {
              "tipo": "item",
              "texto": "11 – Os revestimentos das coberturas em terraço, sem prejuízo do indicado no n.º 6, devem ter uma classe de reação ao fogo mínima igual à indicada no quadro VI abaixo:
                "tipo": "sub_item": [],
                "tipo": "sub_sub_items": [],
                "tabela_refs": ["QUADRO VI", "Reação ao fogo do revestimento das coberturas em terraço",
                "tabela_sumario": "QUADRO VI especifica a reação ao fogo exigida para o revestimento das coberturas em terraço, (...). Já para edifícios com altura superior a '28 m', a exigência aumenta para a classe 'B-s1'.".
                "tabela_sumario_en": "Table VI specifies the required reaction-to-fire classification for terrace roof coverings (...). For buildings with a height greater than '28 m', the requirement is increased to class 'B-s1'.,
                "palavras-chave": {
                  "utilizacao-tipo": {
                    "Portuguese": "A utilização-tipo corresponde à classificação (...) e medidas de proteção aplicáveis...",
                    "English": "The use-type corresponds to the classification (...) and applicable protection measures...",
                    (...)
                  "ID": "PT-Ane0_TitII_CapI_Art3_Ite11", "ID_T": "(...)_Ite11_TabVI"]
            }
          ]
        }
      ]
    }
  ]
}
```

Figura 5: Exemplo do ficheiro JSON enriquecido.

3.5. Criação do Grafo de Regulamentos

Após a etapa de enriquecimento, o conteúdo estruturado é convertido num grafo RDF, permitindo representar de forma clara a estrutura, as relações internas e outra informação relevante dos documentos regulamentares. Esta representação facilita a realização de consultas avançadas, promove a interoperabilidade com outras bases de conhecimento e possibilita a sua utilização em etapas posteriores, como a visualização de dados. Para concretizar esta etapa, Donkers et al. [2] desenvolveram uma ontologia específica, designada *FireBIM Regulation Ontology* (FRO). Esta ontologia leve e modular baseia-se nas recomendações do grupo *Linked Building Data Community* do W3C⁷, incorporando igualmente conceitos já estabelecidos pela ontologia AEC3PO⁸.

A FRO foi especificamente criada para estruturar documentos regulamentares num formato legível por máquina, permitindo a posterior verificação automática da conformidade. Numa fase inicial, esta ontologia não aborda diretamente elementos técnicos específicos (por exemplo, compartimentos corta-fogo ou portas corta-fogo), centrando-se antes na organização lógica e estrutural dos textos regulamentares. A ontologia FRO é constituída por três classes principais que interagem de forma complementar: *fro:Authority*, que identifica a entidade responsável pela publicação e manutenção do regulamento; *fro:DocumentSubdivision*, que representa as diferentes partes internas dos documentos (como capítulos, artigos ou anexos); e *fro:Reference*, que permite estabelecer referências internas ou externas entre documentos.

A classe `fro:DocumentSubdivision` é detalhada através de uma hierarquia de subclasses específicas, tais como `Annex`, `Article`, `Item` e `Block`, sendo esta última subdividida em elementos mais concretos, como `Table`, `Equation`, `Appendix`, `Figure`, `Chapter`, entre outros. As relações estruturais entre estas subclasses são estabelecidas através de propriedades específicas, como `fro:hasAnnex`, `fro:hasTable` ou `fro:hasParagraph`.

Com base nesta ontologia, o ficheiro JSON enriquecido obtido nas etapas anteriores é automaticamente analisado de modo a identificar e extrair a sua estrutura subjacente. Este processo gera automaticamente os indivíduos associados aos elementos concretos do documento, dando origem ao grafo RDF final (Figura 6). Durante esta conversão, são utilizados os identificadores (IDs), a informação multilingue e o conteúdo textual original, garantindo uma representação fidedigna e completa dos dados extraídos. O resultado deste procedimento é exportado em formato Turtle (.ttl), permitindo a realização de consultas avançadas e visualizações através de ferramentas especializadas. Deste modo, o grafo RDF resultante constitui uma infraestrutura semântica robusta, adequada à execução de consultas semânticas complexas.

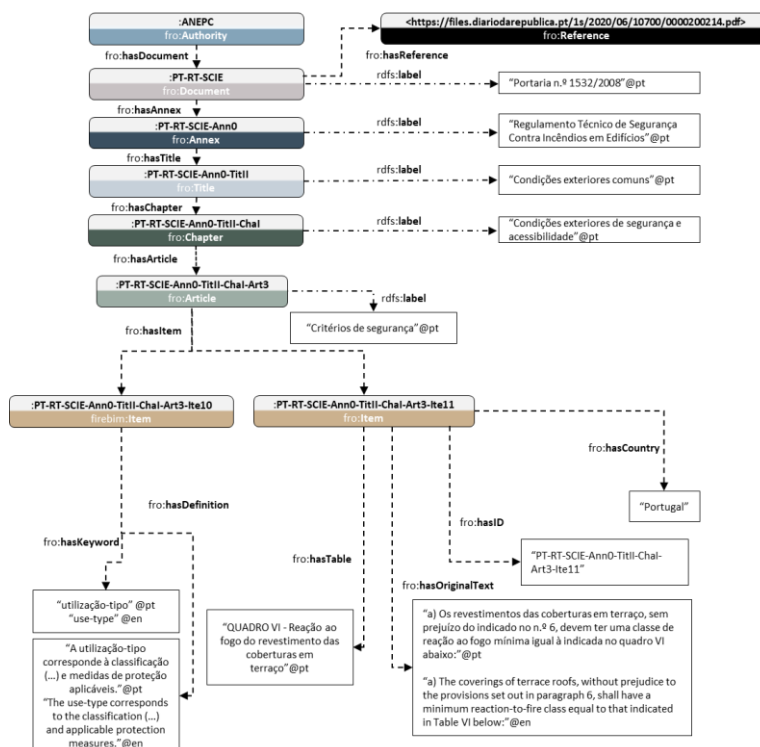


Figura 6: Detalhe do grafo RDF resultante dos regulamentos portugueses.

3.6. Visualização do Grafo de Regulamentos

A fase final da metodologia consiste na visualização interativa dos grafos de conhecimento gerados. Com base no grafo RDF extraído do regulamento em análise, esta etapa permite uma exploração clara, hierárquica e semanticamente enriquecida da informação. Concebida como uma ferramenta de apoio à visualização final, a interface consolida os elementos mais relevantes do documento processado num único ambiente coerente.

A interface HTML encontra-se organizada em áreas funcionais distintas. No canto superior esquerdo, um painel de pesquisa permite aos utilizadores localizar rapidamente termos, artigos ou expressões específicas. Imediatamente abaixo, o painel de palavras-chave e definições

apresenta os principais termos identificados no regulamento, juntamente com as respetivas definições, tanto na língua original do documento como em inglês. Esta informação, proveniente do grafo RDF, encontra-se ligada ao conteúdo apresentado na área principal, permitindo o acesso às definições completas através da seleção dos termos destacados.

O índice do documento, localizado no canto inferior esquerdo, é gerado automaticamente com base na estrutura hierárquica do grafo RDF e apresenta apenas os elementos de nível superior, nomeadamente Título, Capítulo, Secção e Artigo. Apresentado num formato compacto, este índice permite uma navegação direta pelas principais secções do regulamento, facilitando o acesso ao seu conteúdo estrutural. No centro da interface encontra-se a área principal de visualização, onde é apresentado o conteúdo textual do regulamento, enriquecido com ligações semânticas. Esta área inclui igualmente as tabelas extraídas automaticamente na Etapa 3.2, preservadas na sua disposição original e apresentadas como elementos informativos e complementares. No canto superior direito, um menu de visualização permite ajustar parâmetros como o modo escuro, a numeração dos artigos e o nível de detalhe apresentado. Embora ofereça funcionalidades simples, este menu melhora a experiência de leitura e permite adaptar a apresentação do conteúdo às preferências do utilizador.

Em síntese, a principal função desta interface (Figura 7) é servir como uma camada consolidada de consulta e como ferramenta de apoio à validação final. Embora os dados já tenham sido verificados em etapas anteriores, esta visualização integrada permite identificar de forma mais eficaz eventuais inconsistências ou omissões. Caso sejam detetados problemas, a etapa de processamento correspondente deverá ser repetida ou o ficheiro JSON gerado na Etapa 3.4 deverá ser corrigido manualmente. Para além da validação, a publicação desta interface sob a forma de páginas web interativas, associadas a diferentes regulamentos, permitiria igualmente comparar definições de termos semelhantes entre países, acrescentando valor prático à solução proposta. Importa referir que esta interface não foi concebida para integração direta com outras ferramentas.

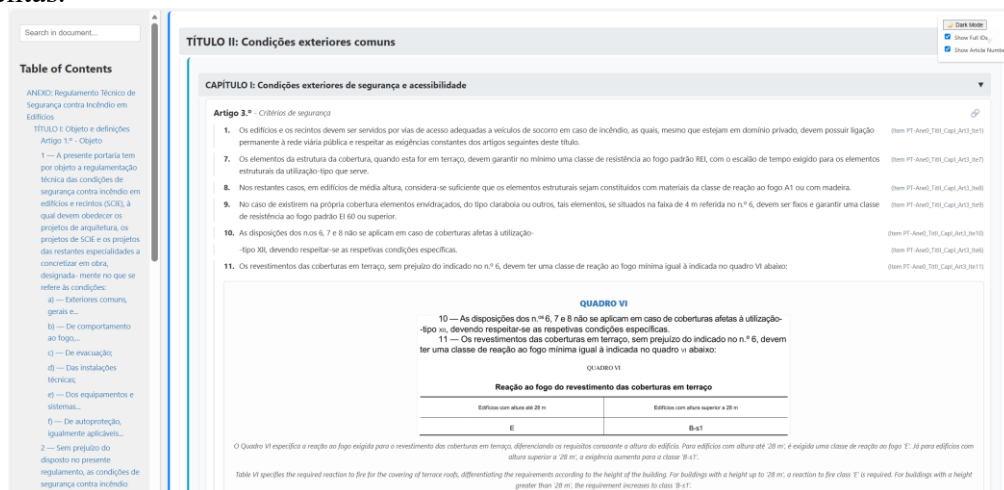


Figura 7: Interface HTML interativa para visualização e validação do documento processado.

4. Conclusões

A eficácia da verificação automática da conformidade em regulamentos de segurança contra incêndio depende da capacidade de converter estes documentos em formatos estruturados e legíveis por máquina. Enquanto a conversão manual é precisa, é morosa e propensa a

inconsistências, e as soluções totalmente automatizadas continuam a enfrentar dificuldades perante a complexidade dos textos regulamentares. Este contraste evidencia a necessidade de metodologias que aumentem a eficiência sem comprometer a fiabilidade. Neste contexto, este estudo propõe uma metodologia inovadora que combina a extração automática de informação baseada em LLMs com validação especializada contínua. A metodologia foi aplicada a regulamentos portugueses de segurança contra incêndio, incluindo documentos extensos e estruturalmente complexos, demonstrando elevada eficácia na captura, estruturação e enriquecimento semântico do conteúdo. Cada documento foi processado em cerca de 1,5 horas, com um custo inferior a 1 €, representando uma melhoria substancial face a abordagens manuais tradicionais. Estes resultados sublinham a importância da digitalização documental e do equilíbrio entre automação e validação humana. Apesar dos resultados promissores, foram identificadas limitações relacionadas com a qualidade da ontologia e a necessidade pontual de ajustes manuais. Como trabalho futuro, prevê-se a criação automática das regras em SHACL Shapes e a validação final do modelo BIM face a estas regras. A metodologia será ainda estendida a regulamentos de diferentes países, com características distintas.

Agradecimentos

Este trabalho foi financiado pelo Programa NORTE 2030 através do Fundo Europeu de Desenvolvimento Regional (FEDER), no âmbito do projeto FireBIM, referência NORTE2030-FEDER-00358400. Este trabalho foi ainda apoiado por: UID/04708/2025 e <https://doi.org/10.54499/UID/04708/2025>, do CONSTRUCT - Instituto de I&D em Estruturas e Construções - financiado pela Fundação para a Ciência e a Tecnologia, I.P./MECI, através dos fundos nacionais.

Referências

- [1] R. K. Chaudhary *et al.*, “Evaluation of anticipated post-fire repair cost for resilient design of composite slab panels,” *Journal of Building Engineering*, vol. 52, p. 104460, 2022, doi: <https://doi.org/10.1016/j.jobe.2022.104460>.
- [2] A. Donkers and E. Petrova, *Converting Fire Safety Regulations to SHACL Shapes Using Natural Language Processing*. 2024.
- [3] N. Nisbet and L. Ma, “Using RASE semantic mark-up for Normative, Definitive and Descriptive knowledge,” *Taylor & Francis*, pp. 1–7, 2023, Accessed: Jul. 15, 2025. [Online]. Available: <https://discovery.ucl.ac.uk/id/eprint/10178378/1/ECPPM-2022-NNa.pdf>
- [4] J. Zhang and N. El-Gohary, “Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking,” *Journal of Computing in Civil Engineering*, vol. 30, p. 141013064441000, Jul. 2013, doi: 10.1061/(ASCE)CP.1943-5487.0000346.
- [5] J. Zhang and N. M. El-Gohary, “Extending Building Information Models Semiautomatically Using Semantic Natural Language Processing Techniques,” *Journal of Computing in Civil Engineering*, vol. 30, no. 5, Sep. 2016, doi: 10.1061/(asce)cp.1943-5487.0000536.
- [6] Costa, G. *et al*, N. Formalization of building codes and regulations in knowledge graphs, in Digital Building Permit conference 2024. Barcelona, Spain, Apr. 2024.
- [7] Al-Turki, D. *et al*. Human-in-the-Loop learning with LLMs for efficient RASE tagging in building compliance regulations, *IEEE Access*, vol. 12, pp. 185291-185306, 2024, doi: 10.1109/ACCESS.2024.3512434.