

Método para Estruturação de Regulamentos Financeiros em SBVR-SE Orientado por Ontologias e LLMs para GRC

Method for Structuring Financial Regulations in SBVR-SE Driven by Ontologies and LLMs for GRC

Anderson dos Santos*, Paulo Sérgio Muniz Silva†

RESUMO

O estudo apresenta um método para transformar textos regulatórios do setor financeiro em regras formais no padrão *Semantics of Business Vocabulary and Business Rules - Structured English (SBVR-SE)*, motivado pelas limitações dos processos atuais de *Governance, Risk and Compliance (GRC)*, que são predominantemente manuais e propensos a erros. A proposta integra fundamentos do SBVR, taxonomias para classificação e padronização das regras de negócio e as capacidades de Large Language Models (LLMs) para interpretação, extração e conversão de textos em linguagem natural para representações estruturadas e auditáveis. O método adota um fluxo em seis etapas, combinando a previsibilidade de técnicas determinísticas com a flexibilidade dos LLMs, preservando rastreabilidade e explicabilidade. A classificação de regras segundo a taxonomia de Witt (2012) assegura alinhamento entre intenção regulatória e representação funcional, favorecendo a automação e o reuso. Embora não implementado, o método se apresenta como base para futuras ferramentas e experimentos, destacando-se pela aplicação de LLMs em conformidade regulatória.

PALAVRAS-CHAVE: SBVR; GRC; FRO; FIBO; Taxonomia; Ontologia; NLP; LLM.

ABSTRACT

This study presents a method for transforming financial sector regulatory texts into formal rules under the Semantics of Business Vocabulary and Business Rules – Structured English (SBVR-SE) standard, motivated by the limitations of current Governance, Risk and Compliance (GRC) processes, which are mainly manual and error-prone. The approach integrates SBVR foundations, taxonomies for business rule classification and standardization, and the capabilities of Large Language Models (LLMs) to interpret, extract, and convert natural language texts into structured, auditable representations. The method follows a six-step workflow, combining the predictability of deterministic techniques with the flexibility of LLMs, ensuring traceability and explainability. Rule classification according to Witt's (2012) taxonomy aligns regulatory intent with functional representation, supporting automation and reuse. Although not yet implemented, the method provides a foundation for future tools and experiments, standing out for its application of LLMs in regulatory compliance.

KEYWORDS: SBVR; GRC; FRO; FIBO; Taxonomy; Ontology; NLP; LLM.

1 INTRODUÇÃO

As instituições financeiras operam em um cenário regulatório cada vez mais denso, dinâmico e suscetível a múltiplas interpretações, o que dificulta a conformidade normativa. Os métodos tradicionais de *Governance, Risk and Compliance (GRC)* permanecem reativos, com auditorias realizadas após a ocorrência de falhas, sendo majoritariamente manuais e sujeitos a erros (Sadiq et al., 2007; Kulkarni et al., 2021; Roychoudhury et al., 2017; Sunkle et al., 2015). Além disso, a interpretação das normas depende de especialistas externos, o que demanda tempo e recursos. A complexidade linguística dos regulamentos, ambiguidade, polissemia e dependência de contexto, torna ainda mais difícil a verificação de conformidade (Hough et al., 2019; Jackson, 2020), exigindo rastreabilidade precisa entre exigências legais e ações organizacionais (Bouzidi et al., 2011). Tecnologias

*  IPT – Instituto de Pesquisas Tecnológicas, São Paulo, SP, Brasil. ✉ adsantos@gmail.com

†  IPT – Instituto de Pesquisas Tecnológicas, São Paulo, SP, Brasil. ✉ paulo.muniz@usp.br

de *Natural Language Processing* (NLP) vêm sendo empregadas para superar essas barreiras, ampliando a capacidade de análise dos textos regulatórios (Khurana et al., 2022).

Nesse contexto as *Controlled Natural Languages* (CNLs), como o *Semantics of Business Vocabulary and Business Rules - Structured English* (SBVR-SE), permitem representar regras com clareza semântica e interoperabilidade entre sistemas (Gruzitis; Barzdins, 2009; OMG, 2019). A transformação de regras de negócio tem sido objeto de diversas abordagens: Abi-Lahoud et al. (2013) propõem um processo semiautomático centrado em especialistas; Joshi e Saha (2020) estruturam as normas do *Code of Federal Regulations* (CFR) em grafos de conhecimento utilizando IA focando em aplicações de consulta; e Haj et al. (2021) desenvolvem um método automático para converter regras explícitas em *Semantics of Business Vocabulary and Business Rules* (SBVR) por meio de NLP, com limitações em estruturas complexas ou ambíguas. Em geral, essas abordagens utilizam algoritmos determinísticos, que apresentam restrições significativas, conforme discutido na seção 3.

Este trabalho descreve um método que utiliza técnicas de NLP utilizando *Large Language Models* (LLMs) para transformar regulamentos do CFR, redigidos em *Natural Language* (NL), para regras e vocabulário estruturados em *Structured English* (SE), conforme a especificação SBVR da *Object Management Group* (OMG). A principal contribuição consiste no desenvolvimento de um método com o formalismo do SBVR para o setor financeiro.

O restante deste artigo está organizado da seguinte forma. A seção 2 apresenta uma breve introdução à SBVR, taxonomia e LLMs no contexto deste trabalho. A seção 3 analisa as abordagens existentes para identificar, extrair e transformar regras de negócio para CNL. A seção 4 detalha o método proposto para a transformação automatizada de regras de negócio para SBVR e a seção 5 conclui o artigo e descreve as direções para novos trabalhos.

2 FUNDAMENTAÇÃO CONCEITUAL

A base conceitual deste trabalho se apoia em três pilares principais: (1) SBVR e CNLs para representar as regras de forma não ambígua; (2) taxonomias para classificar as regras de negócio em relação a sua intenção e padronizar a transformação; (3) e os LLMs para a classificação e transformação dos textos em linguagem natural para SBVR utilizando suas habilidades avançadas de NLP. Esses elementos viabilizam a transformação a partir de NL e a representação estruturada de vocabulários e regras de negócios.

2.1 SBVR E CNL

O SBVR, conforme especificado pela OMG (2019), define uma estrutura formal para representar vocabulários e regras de negócios de modo legível por humanos e processável por máquinas, com base em lógica formal e linguística. Ele organiza termos e definições em dicionários terminológicos e vocabulários de negócios, distinguindo regras de definição e regras comportamentais. Embora não seja uma linguagem controlada, fornece a base para criá-las, como a SBVR-SE apresentada no Anexo C da especificação. Ashfaq e Bajwa (2021) analisam a ambiguidade da NL em requisitos de software e destacam o uso da SBVR-SE pela sua clareza e reutilização de regras. Roychoudhury et al. (2017) utilizam a SBVR-SE como linguagem intermediária para representar regulamentações de forma

compreensível e transformável, enquanto Witt (2012) sugere variações do RuleSpeak¹ para a CNL SBVR. A escolha da SBVR-SE neste trabalho se dá pela sua precisão e alinhamento com a NL.

2.2 TAXONOMIA PARA CLASSIFICAR AS REGRAS DE NEGÓCIO

Não há uma taxonomia amplamente consensual para regras de negócios, Ross (1997) propõe uma classificação baseada em modalidades, como obrigação, proibição, permissão condicional e aconselhamento. A especificação do SBVR (OMG, 2019) distingue regras de definição, que caracterizam conceitos e entidades de negócio, e regras comportamentais, que regulam ações por meio de obrigações, permissões e proibições. Von Halle (2001) contribui com o uso de modelos estruturados para expressar essas regras, utilizando padrões sintáticos que favorecem consistência e automação. Witt (2012) amplia essa abordagem ao classificar regras em definição, dados, atividades e papéis, e distinguir entre permissões e possibilidades, propondo uma taxonomia compatível com o SBVR. Ele também apresenta *templates* para as regras com foco na clareza e uniformidade. Neste trabalho, a taxonomia de Witt (2012) é adotada para identificar padrões de regras e associá-los a *templates* de escrita, promovendo uma formulação estruturada e consistente.

2.3 LLMs NO CONTEXTO DE REGULAMENTOS

Os LLMs, como GPT da OpenAI e Gemini da Google, representam um avanço expressivo em NLP, combinando treinamento em grandes volumes de texto com bilhões de parâmetros. Min et al. (2023) descrevem três paradigmas de uso desses modelos: pré-treinamento com ajuste fino, aprendizado por *prompts* e geração de texto. Sun et al. (2022) complementam essa visão ao mapear sete paradigmas de NLP, destacando a transição para abordagens mais generalistas e eficazes. No contexto jurídico, Anh et al. (2023) mostram que os LLMs superam métodos tradicionais ao lidar com linguagem ambígua e estruturas complexas. Zhong et al. (2020) demonstram que o NLP tem aprimorado tarefas como análise de cláusulas e conformidade, enquanto Goebel et al. (2023) evidenciam isso na prática com a competição *Competition on Legal Information Extraction and Entailment* (COLIEE), onde as equipes aplicam LLMs em tarefas complexas de jurisprudência. Apesar dos avanços, Anh et al. (2023) ressaltam desafios éticos e técnicos, indicando que a aplicação dos LLMs exige atenção à adequação ao domínio e à mitigação de riscos. Nesta pesquisa é proposto o uso de LLMs genéricos, não treinado ou ajustado para um domínio específico para assumir todas as operações de NLP guiado exclusivamente pela engenharia de *prompts*.

3 TRABALHOS RELACIONADOS

Este trabalho introduz um conjunto de padrões para classificar e comparar abordagens de transformação. Baseando-se na noção de padrões como soluções recorrentes a problemas conhecidos (ALEXANDER, 1979), sete padrões foram identificados nos trabalhos estudados: (P1) processamento de documento, que trata da decomposição e preservação da estrutura semântica dos textos; (P2) extração de regras,

¹ RuleSpeak™ é um conjunto de diretrizes para expressar regras de negócios de forma concisa e amigável aos negócios, desenvolvida por Ronald G. Ross, em 1996 (Ver mais em: <https://www.rulespeak.com/en/>. Acesso em: 15/mar./2025).

focado na identificação de trechos que expressam regras de negócio; (P3) extração de conceitos, voltado à identificação e categorização de entidades nos textos; (P4) relacionamento entre conceitos, que lida com sinonímias e generalizações; (P5) desambiguação de conceitos, essencial para a interpretação contextual de termos; (P6) extração de regras; e (P7) transformação de regras para CNL. Essa taxonomia de padrões orienta a avaliação crítica das abordagens existentes e fundamenta o desenvolvimento de uma nova abordagem.

Quadro 1 - Comparação das abordagens que transformam regras de negócio para SBVR

Autores	Resultado	Método	Padrões implementados	Depende de	Nível automação
Abi-Lahoud <i>et al.</i> (2013)	Business vocabulary and rules	Baseado em regras	1-7		Manual
Bajwa <i>et al.</i> (2017)	Business vocabulary and rules	NLP + baseado em regras	2, 3, 6, 7	Diagrama de classes UML	Semi
Chittimalli <i>et al.</i> (2020)	Business vocabulary and rules	NLP + baseado em regras	2-7		Auto
Haj <i>et al.</i> (2021)	Business vocabulary and rules	NLP + baseado em regras	2-7		Auto
Joshi e Saha (2020)	Extração das regras e grafo para QA	NLP + baseado em regras	1-7		Auto
Roychoudhury <i>et al.</i> (2017)	Business vocabulary and rules	NLP + Baseado em regras	2-4, 6, 7	Dicionário	Semi
Skersys <i>et al.</i> (2022)	Business vocabulary and rules	Baseado em regras	1-3, 6, 7	Diagramas BPMN	Semi
Nossa abordagem	Business vocabulary and rules	Modelo probabilístico para NLP e regras	1-7		Auto

Fonte: Autoria própria (2014)

O Quadro 1 compara os trabalhos para transformação de regras de negócio para SBVR obtidos nas fontes de busca ACM, IEEE Explorer, ScienceDirect, Springer Link e Web of Science, no período de 2018 a 2024 buscando pelas palavras-chave: *finance*, *financial*, *natural language processing*, *nlp*, *ontology*, *semantics of business vocabular*. O método de Gava (2022) foi utilizado para a revisão sistemática, além do uso da técnica de

Snowballing (GREENHALGH; PEACOCK, 2005; WOHLIN, 2014). Com base nos padrões apresentados, as abordagens dos autores são discutidas a seguir.

Abi-Lahoud et al. (2013) propõem um protocolo baseado no SBVR para interpretar regulamentações dos EUA. O processo, inteiramente manual (P1-P7), é conduzido por especialistas no assunto, que selecionam e analisam textos legais, identificam modalidades (como obrigações, permissões e proibições), extraem e definem o vocabulário relevante, estruturam regras em SBVR-SE, aplicam a estilização SBVR e formalizam as regras para garantir legibilidade por máquina. A etapa final envolve validação e refinamento das regras SBVR para garantir alinhamento com o texto original. O objetivo é aumentar a precisão e acessibilidade de normas complexas, melhorando a compreensão da conformidade e governança.

A abordagem de Bajwa e Shahzada (2017) para a geração automatizada de restrições *Object Constraint Language* (OCL) a partir de especificações em NL, utiliza o SBVR como um passo intermediário. A abordagem envolve a análise linguística do texto em NL, incluindo marcação de *Parts of Speech* (PoS) e análise baseada em regras, para extrair elementos básicos do SBVR (P2, P3, P6 e P7). Esses elementos são então transformados em sentenças OCL. A abordagem tem algumas limitações, incluindo sua restrição ao domínio do modelo de classe *Unified Modeling Language* (UML), os nomes do vocabulário usados nas restrições devem ser consistentes com os nomes do vocabulário usados no modelo de classe UML e há suporte apenas para certos tipos de sentenças OCL. Os autores não lidam com a representação do documento (P1), desambiguação e generalização de conceitos (P4 e P5).

Chittimalli et al. (2020) propõem um método não supervisionado para extrair vocabulários e regras no formato SBVR a partir de documentos de negócios, adequado para processamento automatizado e verificação em sistemas empresariais. A abordagem é dividida em quatro etapas principais: extração de sentenças (P2), extração de entidades, extração de fatos e mineração de regras (P3, P6 e P7), utilizando técnicas de NLP e analisadores de dependência sintática. Entretanto, a abordagem apresenta limitações importantes, como a ausência de tratamento de correferência (devido às restrições do SpaCy), além de não considerar sinônimos e relações de generalização (P4 e P5). Diferentemente de outras abordagens, utiliza um modelo N-gram para classificar sentenças como regra ou ruído, mas trata cada sentença como independente (P1), o que pode limitar a interpretação de contextos mais amplos.

Haj et al. (2021) propõem um método automático para transformar regras de negócios em NL para o formato SBVR, utilizando técnicas de NLP que incluem análises lexical, sintática e semântica para compreender a estrutura e o significado dos textos (P2). O método foca na extração de um dicionário terminológico (P3, P4 e P5) e na identificação de regras de negócios (P6), que são então mapeadas para elementos SBVR por meio de correspondência de padrões capazes de capturar diferentes estruturas de sentenças, incluindo condições, negações e quantificações (P7). Seu principal objetivo é a construção do dicionário, priorizando regras que impõem restrições ou fornecem informações sobre conceitos, mas excluindo regras relacionadas a processos ou cálculos. Embora eficaz para regras explícitas e bem estruturadas, o método apresenta limitações em lidar com regras complexas ou implícitas que exigem conhecimento de domínio, e trata todas as cláusulas como independentes, ignorando hierarquias (P1).

Joshi e Saha (2020) propõem a conversão de regulamentações do CFR em um formato estruturado e legível por máquinas, utilizando técnicas de NLP e aprendizado profundo para extrair termos, definições e regras (P2, P3, P4, P6), os quais são organizados

em um grafo de conhecimento semanticamente enriquecido (P1). Aplicado ao Título 48 da CFR, esse grafo permite raciocínio automatizado e pode ser usado em sistemas de perguntas e respostas, conformidade e análise regulatória, embora não vise a transformação direta em regras SBVR (P7). Os autores também abordam a ambiguidade terminológica criando um dicionário contextualizado (P5) validado por especialistas. Para isso, combinam técnicas baseadas em regras e expressões regulares com modelos como *Recurrent Neural Network* (RNN), Word2Vec e LLMs. No experimento, o LLM foi pré-treinado com tarefas de linguagem e ajustado com dados regulatórios, alcançando alto desempenho na classificação de cláusulas de obrigação, permissão e proibição.

Roychoudhury et al. (2017) propõem uma abordagem semiautomatizada para transformar texto jurídico para *Structured English* (SE), visando a verificação de conformidade no setor financeiro. A metodologia detecta sentenças e extrai proposições n-árias (P2, P6) por meio da ferramenta ClausIE, seguida pela seleção baseada em um cálculo de "informatividade" - razão entre menções reconhecidas no modelo de domínio e o total de tokens da proposição. Conceitos, definições e sinônimos são identificados com técnicas de aprendizado de máquina e NLP (P3, P4), e as proposições selecionadas são convertidas para SE (P7). A transformação subsequente de SE para SBVR é automática, embora os autores não detalhem o mapeamento. Apesar das limitações, como dependência do ClausIE e a necessidade de os especialistas dominarem a notação SE, a abordagem permite que regras regulatórias sejam estruturadas por especialistas sem conhecimento formal em SBVR. A desambiguação de termos (P5) não é tratada, bem como a dependência das cláusulas (P1).

Skersys et al. (2022) propõem uma metodologia para converter diagramas BPMN em regras SBVR, utilizando lógica deôntica para incorporar obrigações, permissões e proibições. O método identifica os elementos *Business Process Model and Notation* (BPMN) (tarefas, eventos, gateways) e seu mapeamento semântico para conceitos equivalentes no SBVR (P1, P2, P3), seguido pela aplicação de um conjunto de nove transformações específicas, que automatizam a conversão para regras SBVR (P7). Essas transformações consideram desde fluxos de sequência simples até interações complexas entre participantes, mantendo a integridade semântica do modelo original. A abordagem requer a correta estruturação dos diagramas BPMN, incluindo a definição explícita de participantes e o uso padronizado de nomes e condições. Apesar da eficácia observada nos testes com 32 modelos, os autores reconhecem limitações como a validade de construção e interna, a dependência de modelagem manual e a ausência de práticas padronizadas, o que pode comprometer a generalização dos resultados.

Chittimalli et al. (2020) e Haj et al. (2021) destacam-se por incluir em seus métodos a extração automática de dicionários terminológicos diretamente do texto, ao contrário de outras abordagens que dependem de fontes externas ou não especificam como obtêm as definições. A qualidade desses dicionários impacta diretamente a precisão da transformação, sendo aprimorada quando ontologias reconhecidas para o domínio, como sugerido por Ford et al. (2016), são utilizadas. Embora todas as abordagens analisadas compartilhem a característica de transformação baseada em regras e o uso de NLP para decompor e parametrizar elementos do texto, enfrentam limitações quanto à generalização, sendo seus resultados possivelmente não replicáveis em outros contextos. Além disso, a maioria trata os fragmentos textuais como independentes, o que não reflete a estrutura dos textos legais (AGNOLONI; FRANCESCONI, 2011), com exceções notáveis para Joshi e Saha (2020) e Dragoni et al. (2018), que propõem modelos ontológicos estruturais.

Nenhuma das abordagens revisadas disponibiliza publicamente os códigos, dados ou artefatos que possibilitem a reprodução dos experimentos.

4 MÉTODO PARA A TRANSFORMAÇÃO

A extração automática de regras e políticas a partir de documentos regulatórios ainda apresenta limitações relevantes, pois técnicas tradicionais de NLP não capturam as relações semânticas complexas e a estrutura hierárquica dos textos legais (Louis et al., 2023; Anh et al., 2023; Sun et al., 2022). Joshi e Saha (2020) defendem a combinação de múltiplas abordagens para representar adequadamente o conhecimento jurídico. Este trabalho utiliza o protocolo de Abi-Lahoud et al. (2013), que define um processo estruturado com etapas de leitura, identificação e estilização de modalidades, verbos e conceitos, seguido do enriquecimento semântico dos elementos, assegurando precisão terminológica e consistência na formulação de regras no padrão SBVR.

Figura 1 - Exemplo da transformação de uma regra de negócio em NLP para SBVR

Title: Evidence of Service Certification

CS3: (a)(3) If the Secretary certifies that the Commission was served with process, pleadings, or other papers pursuant to paragraph (a)(1) of this section and forwarded these documents to a named party pursuant to paragraph (a)(2) of this section, this certification constitutes evidence of service upon that party. [Party rules | Responsibility rules]

Rule

TS3: (a)(3) The certification by the Secretary must constitute evidence of service upon a named party if the Secretary certifies that the Commission was served with process, pleadings, or other papers pursuant to paragraph (a)(1) of this section and forwarded these documents to a named party pursuant to paragraph (a)(2) of this section. [Party rules | Responsibility rules]

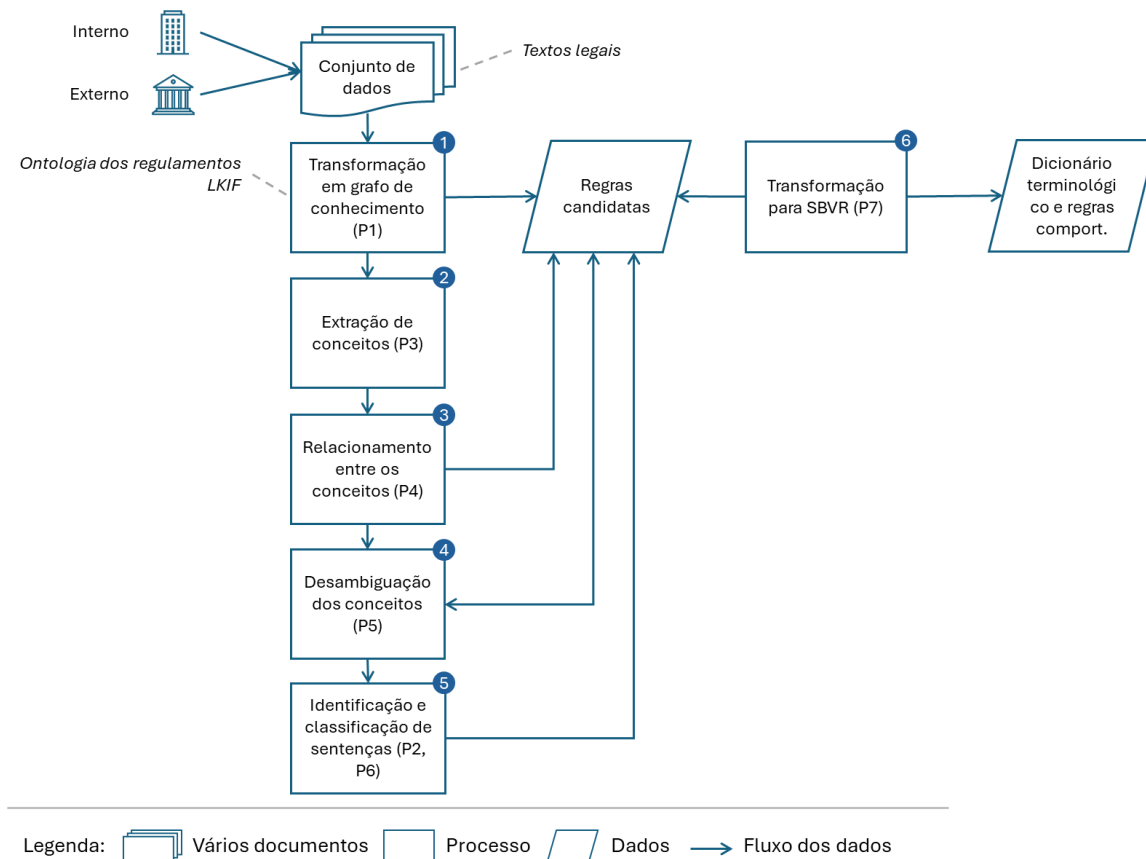
Fonte: Autoria própria (2024)

Um exemplo, obtido pela aplicação manual do protocolo, é apresentado na Figura 1, nele a primeira sentença foi obtida do título 17, capítulo II, parte 275, seção 275.0-2 da CFR² (parágrafo a, subparágrafo 3). Esse trecho foi classificado, segundo a taxonomia de Witt (2012), como uma regra de negócio operativa do tipo papel, subtipo responsabilidade. As decorações no texto indicam os elementos identificados: duplo sublinhado são nomes, sublinhado simples termos, em itálico e azul os símbolos verbais e em normal e laranja as palavras-chave. Cada um dos elementos, incluindo a própria declaração, são relacionados em um grafo de conhecimento, juntamente com os metadados de classificação (tipo do elemento, classe na taxonomia etc.).

Além da identificação da regra de negócio, para GRC é necessário preservar a rastreabilidade entre a regra candidata (*candidate statement* - CS) e a sua transformada (*transformed statement* - TS), bem como todos os resultados intermediários no processo de transformação (explicabilidade).

² <https://www.ecfr.gov/current/title-17/chapter-II/part-275/section-275.0-2>. Acesso em: 09/ago./2025

Figura 2 - Método para o processo de transformação de NL para SBVR



Fonte: Autoria própria (2024)

Para obter o resultado da Figura 1, ao protocolo foram incorporadas etapas adicionais com base nos trabalhos de Ashfaq e Bajwa (2021), Omrane et al. (2011) e Haj et al. (2021), que aplicam a anotação semântica e padrões para a identificação e mapeamento de declarações para SBVR com o objetivo de associar metadados aos elementos para reduzir ambiguidades e aproximar a NL da representação formal, além de permitir que especialistas compreendam melhor os elementos e as transformações aplicadas. O método proposto é organizado em seis etapas principais, conforme ilustrado na Figura 2. As etapas 1 e 2 concentram-se na identificação e extração dos elementos relevantes do texto legal, tais como termos e regras operativas. As etapas 3 a 5 classificam as declarações segundo a taxonomia de Witt (2012), de forma a prover a intenção do elemento e o *template* para sua transformação. A etapa 6 transforma os elementos anotados para o formato SBVR-SE, tratando frases longas, múltiplas cláusulas e ambiguidades terminológicas (Wyner; Peters, 2011), bem como a identificação de regras atômica (Witt, 2012 e Haarst, 2013).

Enquanto abordagens anteriores, como a de Haj et al. (2021), utilizam algoritmos determinísticos baseados em informações linguísticas derivadas de técnicas de NLP, como variações de PoS, para classificar e transformar elementos em estruturas compatíveis com uma CNL, este trabalho propõe uma estratégia baseada em LLMs guiados por *prompts*. Cada etapa do método é executada independentemente com base nas capacidades de classificação e geração de texto dos LLMs e de um contexto global compartilhado. A divisão do método em etapas coesas simplifica as instruções buscando reduzir a propensão a erros

e aumentar a consistência dos resultados, integrando a previsibilidade dos métodos determinísticos à flexibilidade e adaptabilidade dos LLMs (He, 2024; Sun et al., 2024).

5 CONCLUSÃO E TRABALHOS FUTUROS

A motivação deste trabalho são as limitações dos processos de GRC, especialmente no setor financeiro, quanto à automação e formalização de políticas e regulamentos. Embora o uso do formalismo do SBVR possa trazer benefícios significativos, são escassos os trabalhos que o integram com as exigências específicas de GRC, tais como rastreabilidade e explicabilidade. A combinação do SBVR com novas tecnologias de IA abre novas possibilidades para automação, especialmente nas tarefas complexas de identificar e transformar regulamentações financeiras em regras formais. LLMs contribuem com a extração e estruturação dessas regras por meio da sua capacidade de interpretação linguística, mas apresentam limitações importantes, como vieses, imprecisões e falta de explicabilidade, que exigem sua aplicação em fluxos controlados e auditáveis. A classificação dos elementos segundo uma taxonomia contribui para alinhar a intenção regulatória com seu significado funcional, padronizando representações e promovendo o reuso e a automação das regras de negócio.

Este trabalho focou em estabelecer um método para construção de soluções automatizadas que contribui para modernizar as propostas de trabalhos anteriores, inclui uma perspectiva prática da utilização de LLMs e serve de ponto de partida para futuras implementações. Embora não testado, o uso de LLMs é promissor como indicam alguns trabalhos (Chen et al., 2023, Anh et al., 2023 e Yu et al., 2023).

Há diversas direções para continuidade deste trabalho, entre elas a construção de ferramentas que implementem este método, experimentos que afirmem os seus resultados e a forma como os especialistas e máquinas interagirão com esses resultados para validá-los e utilizá-los. Um problema imediato e essencial para a validação de qualquer solução é a escassez de conjuntos de dados ouro (Wyner; Peters, 2011). A construção desses conjuntos de dados é dispendiosa e necessita de especialistas no domínio e SBVR. Para endereçar esse problema a construção de dados sintéticos (ENDRES et al., 2022) que combinem técnicas de ajuste-fino, como *direct-Reinforcement Learning from AI Feedback* (LEE et al., 2023), e um feedback de equivalência semântica gerado por IA são alternativas a serem exploradas.

Agradecimentos

Agradeço ao Instituto de Pesquisa Tecnológicas do Estado de São Paulo (IPT) onde contei com a colaboração dos professores para o desenvolvimento deste trabalho.

Conflito de interesse

Não há conflito de interesse.

REFERÊNCIAS

ABI-LAHOUD, Elie et al. Interpreting Regulations with SBVR, in. Collaboration Technologies and Systems (CTS), **IEEE**, p. 439-445, Maio 2013.

AGNOLONI, T.; FRANCESCONI, E.. Modelling semantic profiles in legislative documents for enhanced norm accessibility. **Proceedings Of The 13Th International Conference On Artificial Intelligence And Law**, [S.l.], p. 111-115, 6 jun. 2011. ACM.

ALEXANDER, Christopher. **The timeless way of building**. New York: Oxford university press, 1979.

ANH, Dang Hoang et al. The impact of large language modeling on natural language processing in legal texts: A comprehensive survey. In: **2023 15th International Conference on Knowledge and Systems Engineering (KSE)**. IEEE, 2023. p. 1-7.

ASHFAQ, Fariha; BAJWA, Imran Sarwar. Natural language ambiguity resolution by intelligent semantic annotation of software requirements. **Automated Software Engineering**, [S.l.], v. 28, n. 2, p. 13-13, 21 jul. 2021. Springer Science and Business Media LLC.

BAJWA, Imran Sarwar; SHAHZADA, Muhammad Anwar. Automated generation of OCL constraints: NL based approach vs pattern based approach. **Mehran University Research Journal of Engineering & Technology**, [S.l.], v. 36, n. 2, p. 243-254, abr. 2017.

BOUZIDI, Khalil Riad et al. An Ontological Approach for Modeling Technical Standards for Compliance Checking. **Lecture Notes In Computer Science**, [S.l.], p. 244-249, 2011. Springer Berlin Heidelberg.

CHEN, Kua et al. Automated Domain Modeling with Large Language Models: A Comparative Study. In: 2023 ACM/IEEE 26th International Conference on Model Driven Engineering Languages and Systems (MODELS). **IEEE**, p. 162-172, 2023.

CHITTIMALLI, Pavan Kumar et al. An Approach to Mine SBVR Vocabularies and Rules from Business Documents. **Proceedings Of The 13Th Innovations In Software Engineering Conference On Formerly Known As India Software Engineering Conference**, [S.l.], p. 1-11, 27 fev. 2020. ACM.

DRAGONI, Mauro et al. Combining Natural Language Processing Approaches for Rule Extraction from Legal Documents. **Lecture Notes In Computer Science**, [S.l.], p. 287-300, 2018. Springer International Publishing.

ENDRES, Markus et al. Synthetic data generation: A comparative study. In: **Proceedings of the 26th international database engineered applications symposium**. 2022. p. 94-102.

FORD, Reginald et al. Automating Financial Regulatory Compliance Using Ontology+Rules and Sunflower. **Proceedings Of The 12Th International Conference On Semantic Systems**, [S.l.], p. 113-120, 12 set. 2016. ACM.

GAVA, Vagner Luiz. **Metodologia de pesquisa científica**: Aula 01. São Paulo: Instituto de Pesquisas Tecnológicas do Estado de São Paulo, 2022.

GOEBEL, Randy *et al.* Summary of the Competition on Legal Information, Extraction/Entailment (COLIEE) 2023. **Proceedings Of The Nineteenth International Conference On Artificial Intelligence And Law**, [S.I.], p. 472-480, 19 jun. 2023. ACM.

GREENHALGH, Trisha; PEACOCK, Richard. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. **BMJ**, [S.I.], v. 331, n. 7524, p. 1064-1065, 17 out. 2005. BMJ.

GRUZITIS, Normunds; BARZDINS, Guntis. Polysemy in controlled natural language texts. In: **International Workshop on Controlled Natural Language**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 102-120.

HAJ, Abdellatif *et al.* The Semantic of Business Vocabulary and Business Rules: an automatic generation from textual statements. **IEEE Access**, [S.I.], v. 9, p. 56506-56522, 2021. Institute of Electrical and Electronics Engineers (IEEE).

HAARST, R. V. **SBVR Made Easy**: Business Vocabulary and Rules as a Critical Asset. Amsterdam: Wardy Poelstra, 2013.

HE, Qianyu *et al.* From Complex to Simple: enhancing multi-constraint complex instruction following ability of large language models. **[S.I.: s.n.]**, p. 1-18, jun. 2024.

HOUGH, Alexander R.; GLUCK, Kevin A. The understanding problem in cognitive science. **Advances in Cognitive Systems**, [S.I.], v. 8, p. 13-32, dez. 2019.

JACKSON, Philip. Understanding understanding and ambiguity in natural language. **Procedia Computer Science**, [S.I.], v. 169, p. 209-225, 2020. Elsevier BV.

JOSHI, Karuna Pande; SAHA, Srishty. A semantically rich framework for knowledge representation of code of federal regulations. **Digital Government: Research and Practice**, v. 1, n. 3, p. 1-17, 2020.

KHURANA, Diksha *et al.* Natural language processing: state of the art, current trends and challenges. **Multimedia Tools And Applications**, [S.I.], v. 82, n. 3, p. 3713-3744, 14 jul. 2022. Springer Science and Business Media LLC.

KULKARNI, Vinay *et al.* Toward automated regulatory compliance. **CSI Transactions On Ict**, [S.I.], v. 9, n. 2, p. 95-104, jun. 2021. Springer Science and Business Media LLC.

LEE, Harrison *et al.* RLAIIF vs. RLHF: Scaling reinforcement learning from human feedback with ai feedback. **arXiv preprint arXiv:2309.00267**, 2023.

LOUIS, Antoine *et al.* Finding the Law: enhancing statutory article retrieval via graph neural networks. **Proceedings Of The 17Th Conference Of The European Chapter Of The Association For Computational Linguistics**, [S.I.], p. 2753-2768, 2023. Association for Computational Linguistics.

MIN, Bonan *et al.* Recent Advances in Natural Language Processing via Large Pre-trained Language Models: a survey. **ACM Computing Surveys**, [S.I.], v. 56, n. 2, p. 1-40, 14 set. 2023. Association for Computing Machinery (ACM).

OMG - Object Management Group. **Semantics of Business Vocabulary and Business Rules Version 1.5**. 2019. Disponível em: <https://www.omg.org/spec/SBVR/1.5/PDF>. Acesso em: 22 out. 2023.

ROYCHOUDHURY, Suman *et al.* From natural language to SBVR model authoring using structured English for compliance checking. *In: 2017 IEEE 21st International Enterprise Distributed Object Computing Conference (EDOC)*. **IEEE**, p. 73-78, 2017.

ROSS, Ronald G. Classification of Rule Types. *In: 1997. The Business Rule Book: classifying, defining and modeling rules: Ross method, version 4.0*. Pennsylvania: Database Research Group, 1997. p. 1-394.

SADIQ, Shazia; GOVERNATORI, Guido; NAMIRI, Kioumars. Modeling control objectives for business process compliance. *In: International conference on business process management*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 149-164.

SHANKAR, Shreya *et al.* Who Validates the Validators? Aligning LLM-Assisted Evaluation of LLM Outputs with Human Preferences. **Proceedings Of The 37Th Annual Acm Symposium On User Interface Software And Technology**, [S.I.], p. 1-14, 11 out. 2024. ACM.

SKERSYS, Tomas *et al.* Transforming BPMN Processes to SBVR Process Rules with Deontic Modalities. **Applied Sciences**, [S.I.], v. 12, n. 18, p. 8976, 7 set. 2022. MDPI AG.

SUN, Haoran *et al.* Conifer: improving complex constrained instruction-following ability of large language models. **[S.N.]**, [S.I.], p. 1-16, abr. 2024. ArXiv.

SUN, Tian-Xiang *et al.* Paradigm Shift in Natural Language Processing. **Machine Intelligence Research**, [S.I.], v. 19, n. 3, p. 169-183, 28 maio 2022. Springer Science and Business Media LLC.

SUNKLE, Sagar *et al.* Toward Better Mapping between Regulations and Operations of Enterprises Using Vocabularies and Semantic Similarity. **Complex Systems Informatics And Modeling Quarterly**, [S.I.], n. 5, p. 39-60, 29 dez. 2015.

VON HALLE, Barbara. **Business rules applied: building better systems using the business rules approach**. [S.I.]: Wiley Publishing, oct. 2001.

WITT, Graham. **Writing effective business rules**. Elsevier, 2012.

WOHLIN, Claes. Guidelines for snowballing in systematic literature studies and a replication in software engineering. **Proceedings Of The 18Th International Conference On Evaluation And Assessment In Software Engineering**, [S.I.], p. 1-10, 13 maio 2014. ACM.

WYNER, Adam; PETERS, Wim. On rule extraction from regulations. *In: Legal knowledge and information systems*. **IOS Press**, [S.l.], v. 235, p. 113-122, 2011.

YU, Fangyi *et al.* Exploring the Effectiveness of Prompt Engineering for Legal Reasoning Tasks. **Findings Of The Association For Computational Linguistics**: ACL 2023, [S.l.], p. 13582-13596, 2023. Association for Computational Linguistics.

ZHONG, Haoxi *et al.* How Does NLP Benefit Legal System: a summary of legal artificial intelligence. **[S.N.]**, [S.l.], p. 1-13, maio 2020. ArXiv.