

# **Transformação automatizada de regulamentos financeiros para SBVR:**

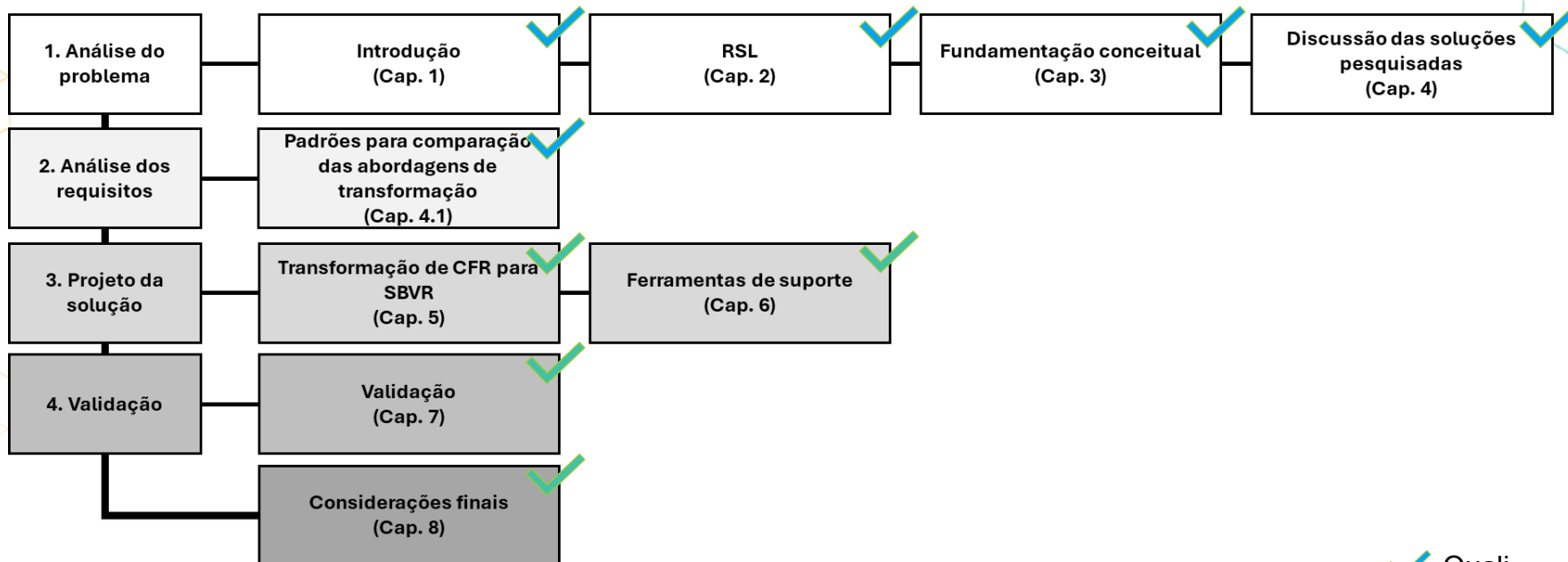
Uma abordagem baseada em processamento de linguagem natural e ontologias

Anderson Santos (orientado)  
Prof. Dr. Paulo Sergio Muniz Silva (orientador)

Mestrado Ciência da Computação  
2025-05-22

1.	Revisão da quali	5 min
2.	Experimento	15 min
3.	Resultados	8 min
4.	Discussão e conclusão	5 min
5.	Trabalhos futuros	2 min

Figura 2 - Organização do trabalho



Fonte: Elaborado pelo autor

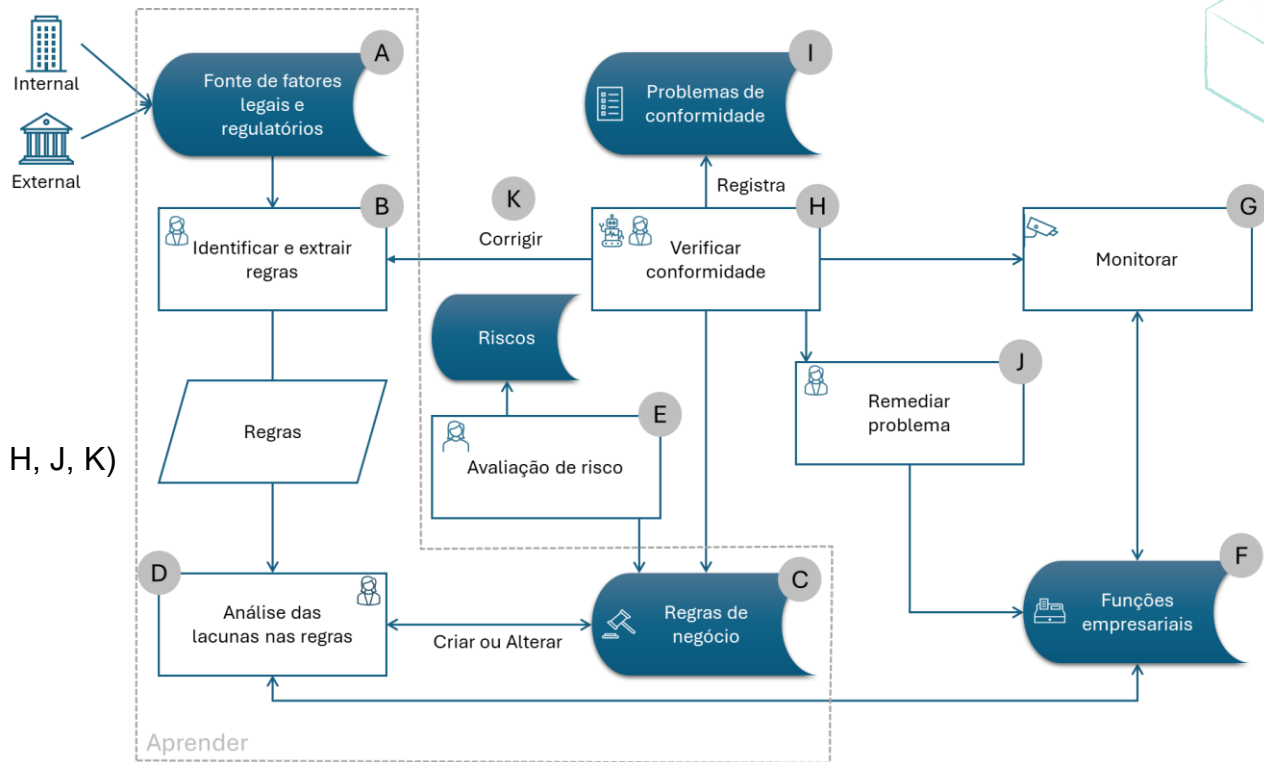
✓ Quali  
✓ Defesa

Figura 5 - Fluxo conceitual para conformidade, baseado no livro vermelho GRC (OCEG, 2023) combinado com processo de três etapas para gerenciar a conformidade (RACZ *et al.*, 2010)

Identificação e extração (A, B, D)


Identificação de riscos (E)

Operacionalização das regras (D, G, H, J, K)



Legenda: Dados armazenados Processo Dados Fluxo dos dados

Quadro 6 - Comparação das abordagens que transformam regras de negócio para SBVR

<b>Autores</b>	<b>Resultado</b>	<b>Método</b>	<b>Padrões implementados</b>	<b>Depende de</b>	<b>Nível automação</b>
Abi-Lahoud et al. (2013)	Business vocabulary and rules	Baseado em regras	1-7		Manual
Bajwa et al. (2017)	Business vocabulary and rules	NLP + baseado em regras	2, 3, 6, 7	Diagrama de classes UML	Semi
Chittimalli et al. (2020)	Business vocabulary and rules	NLP + baseado em regras	2-7		Auto
Haj et al. (2021a)	Business vocabulary and rules	NLP + baseado em regras	2-7		Auto
Joshi et al. (2020)	Extração das regras e grafo para QA	NLP + baseado em regras	1-7		Auto
Polonio (2018)	Business vocabulary and rules	Baseado em regras	1-7	BPMN, UML	Manual
Roychoudhury et al. (2017)	Business vocabulary and rules	NLP + Baseado em regras	2-4, 6, 7	Dicionário	Semi
Skersys et al. (2022)	Business vocabulary and rules	Baseado em regras	1-3, 6, 7	Diagramas BPMN	Semi
CFR2SBVR	Business vocabulary and rules	NLP (Modelo probabilístico) 	1-7	FIBO	Auto

Fonte: Elaborado pelo autor

## Pergunta de pesquisa

*Como a aplicação de técnicas avançadas de processamento em linguagem natural e ontologias, como a FIBO, podem aprimorar a transformação automatizada de regulamentos financeiros em regras e vocabulário estruturados, conforme a especificação SBVR?*

*Propor, implementar e validar um método automatizado que empregue técnicas avançadas de NLP para transformar regulamentos Code of Federal Regulations (CFR), escritos em NL, para regras e vocabulário em inglês estruturado (SE) conforme a especificação SBVR, abreviada para CFR2SBVR.*

*A aplicação de um processo automatizado baseado em técnicas avançadas de NLP com um modelo de IA probabilístico pré-treinado, o uso de grafo de conhecimento para representar textos legais e o suporte da FIBO para o vocabulário deverão resultar em uma **melhora significativa** na precisão e eficiência da transformação.*

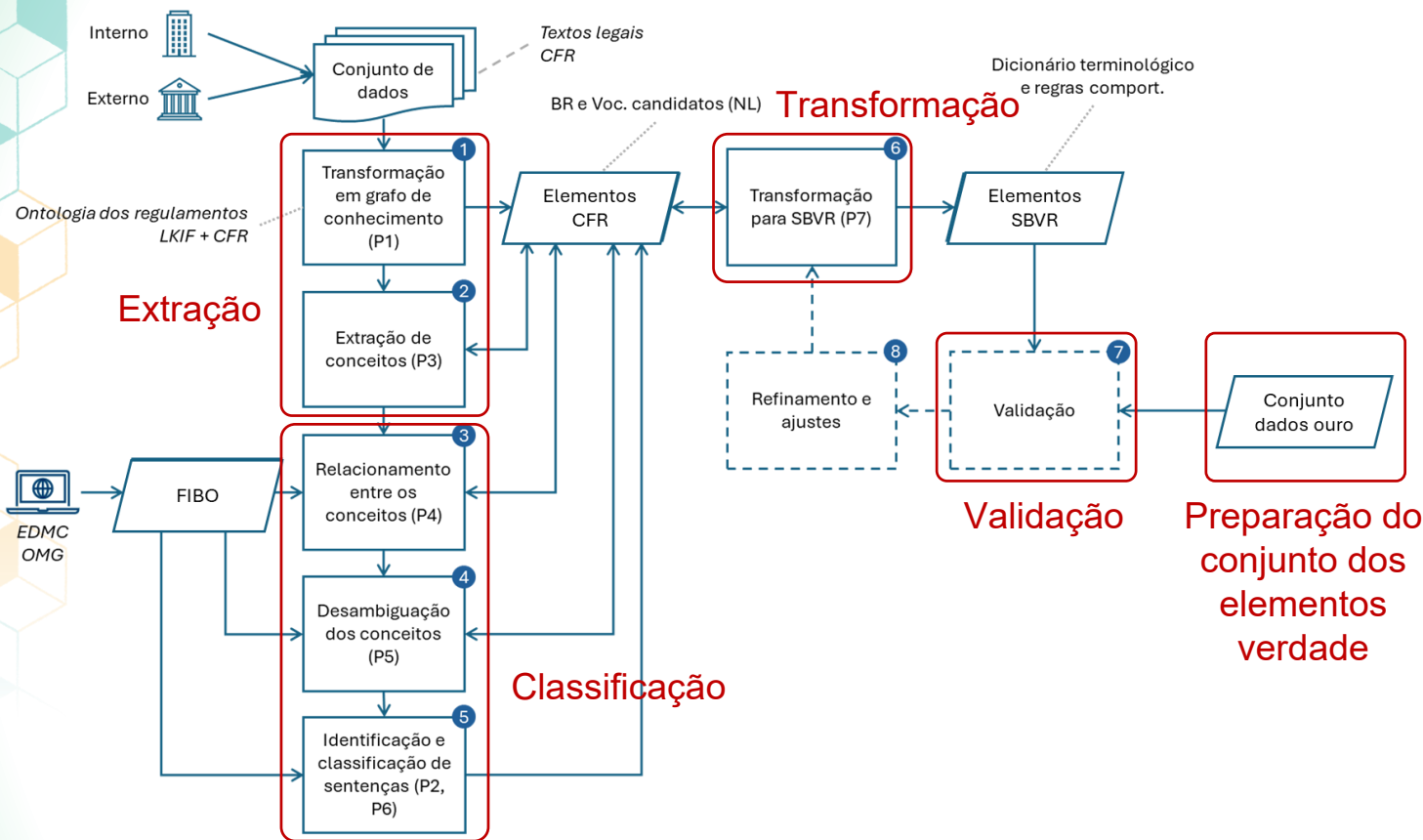


## Contribuições

1. A identificação do **estado da arte** para abordagens de transformação de regras em NL para linguagem controlada.
2. A **implementação** de um método automatizado para transformar textos de regras de negócio em NL para regras em SBVR.
3. A **validação** deste método é feito **por um experimento**, aplicando métodos estatísticos quantitativos para compará-la com um conjunto de dados de referência (padrão-ouro) gerados por esta pesquisa e métricas de similaridade semântica.

1.	Revisão da quali	5 min
2.	<b>Experimento</b>	<b>15 min</b>
3.	Resultados	8 min
4.	Discussão e conclusão	5 min
5.	Trabalhos futuros	2 min

Figura 7- Arcabouço para o processo de transformação de NL para SBVR.



Baseado no protocolo de Abi-Lahoud et al. (2013) combinado com Haj et al. (2021a), Joshi et al. (2020) e Omrane et al. (2011)



# Conjunto de dados ouro

Baseado no protocolo de Abi-lahoud (2013) - Interpreting Regulations with SBVR

Title 17 / Chapter II / Part 275

Previous / Next / 1

Table of Contents

Details

Print/PDF

Display Options

Subscribe

Timeline

Go to Date

Compare Dates

Published Edition

## § 275.0-5 Procedure with respect to applications and other matters.

The procedure hereinbelow set forth will be followed with respect to any proceeding initiated by the filing of an application, or upon the Commission's own motion, pursuant to any section of the Act or any rule of regulation thereunder, unless in the particular case a different procedure is provided:

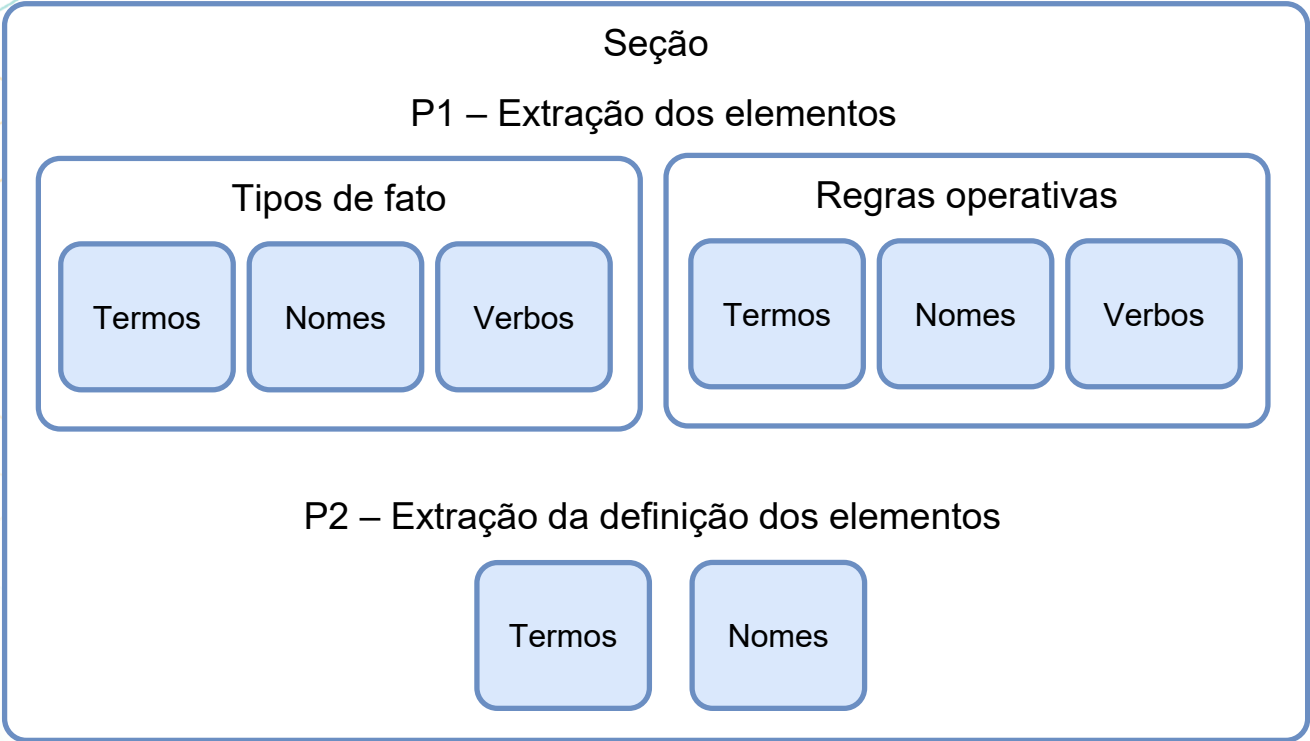
- (a) Notice of the initiation of the proceeding will be published in the FEDERAL REGISTER and will indicate the earliest date upon which an order disposing of the matter may be entered. The notice will also provide that any interested person may, within the period of time specified therein, submit to the Commission in writing any facts bearing upon the desirability of a hearing on the matter and may request that a hearing be held, stating his reasons therefor and the nature of his interest in the matter.
- (b) An order disposing of the matter will be issued as of course following the expiration of the period of time referred to in paragraph (a) of this section, unless the Commission thereafter orders a hearing on the matter.
- (c) The Commission will order a hearing on the matter, if it appears that a hearing is necessary or appropriate in the public interest or for the protection of investors,
  - (1) upon the request of any interested person or
  - (2) upon its own motion.

Regra atômica

```
1 documents_true_table.json X
2 code > data > {} documents_true_table.json > {} § 275.0-5_P1[true_table > {} content > {} elements > {} 3
515 "§ 275.0-5_P1[true_table": {
518   "content": {
521     "elements": [
613       {
651         "id": 4,
652         "title": "Commission orders a hearing.",
653         "statement": "The Commission will order a hearing on the matter, if it appears",
654         "terms": [
656           {
658             "term": "Motion",
659             "classification": "Common Noun"
660           },
661           {
662             "term": "Commission",
663             "classification": "Proper Noun"
664           },
665           {
666             "term": "Hearing",
667             "classification": "Common Noun"
668           },
669           {
670             "term": "Matter",
671             "classification": "Common Noun"
672           },
673         ]
685       },
686       "verb_symbols": [
687         "will order",
688         "appears",
689         "is necessary",
690         "is appropriate"
691       ],
692       "classification": "Operative Rule",
693       "sources": [
694         "(c)"
695       ]
696     ]
  }
}
```



# Estratégia de extração



## Manifesto das Regras de Negócio

[...] 3.1. As regras constroem-se sobre factos, e os factos sobre conceitos que são expressos por termos. [...]

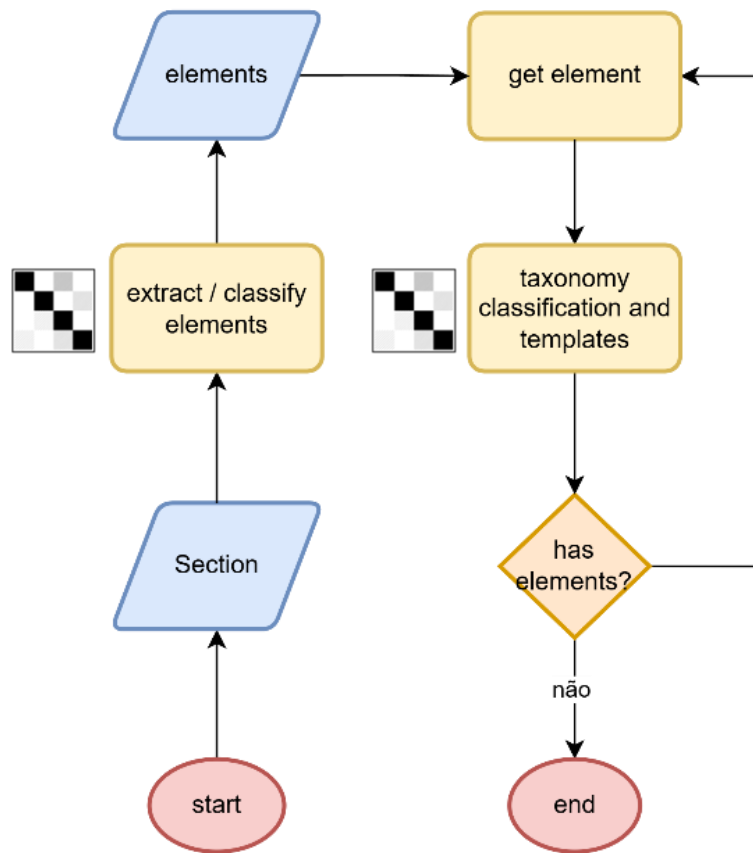


Extração



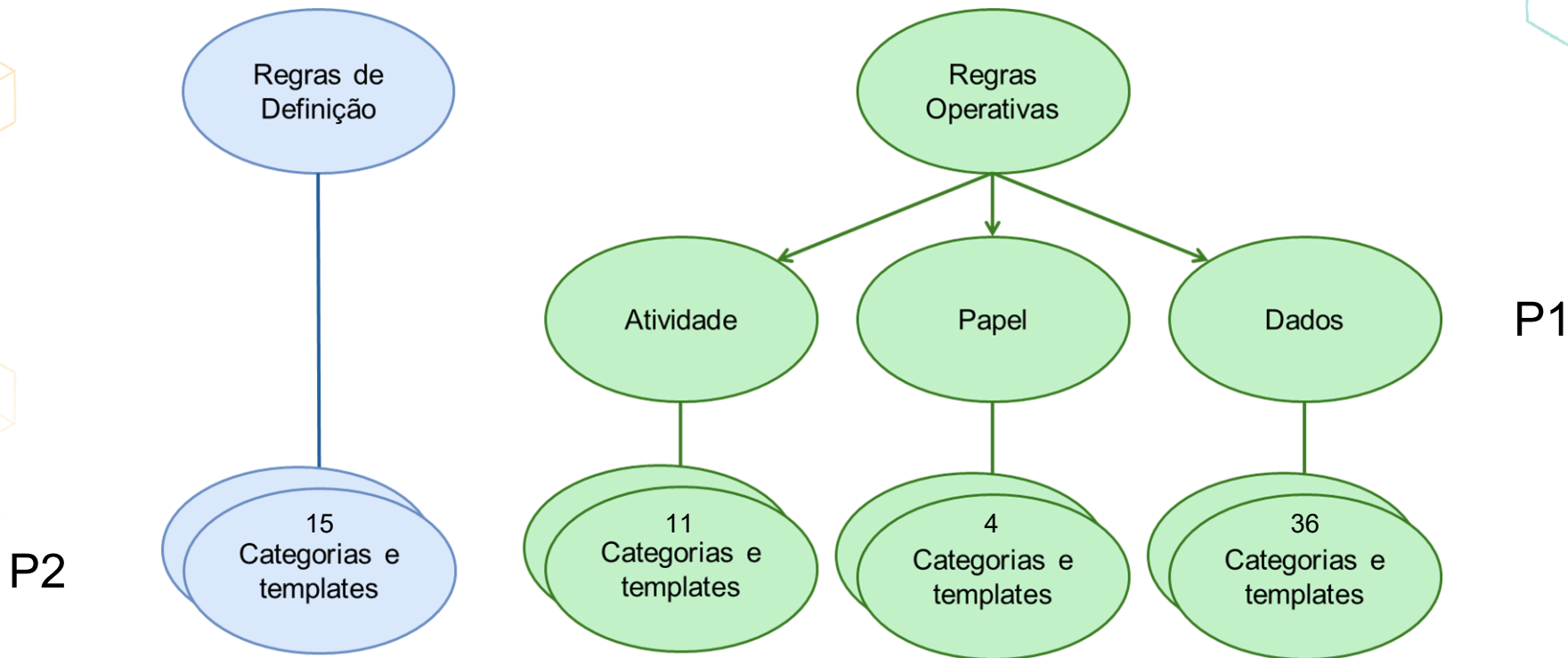
Classificação

Figura 17 - Algoritmo “semantic annotation”



Fonte: Elaborado pelo autor

Figura 6 - Taxonomia de Witt (2012)



Fonte: Elaborado pelo autor

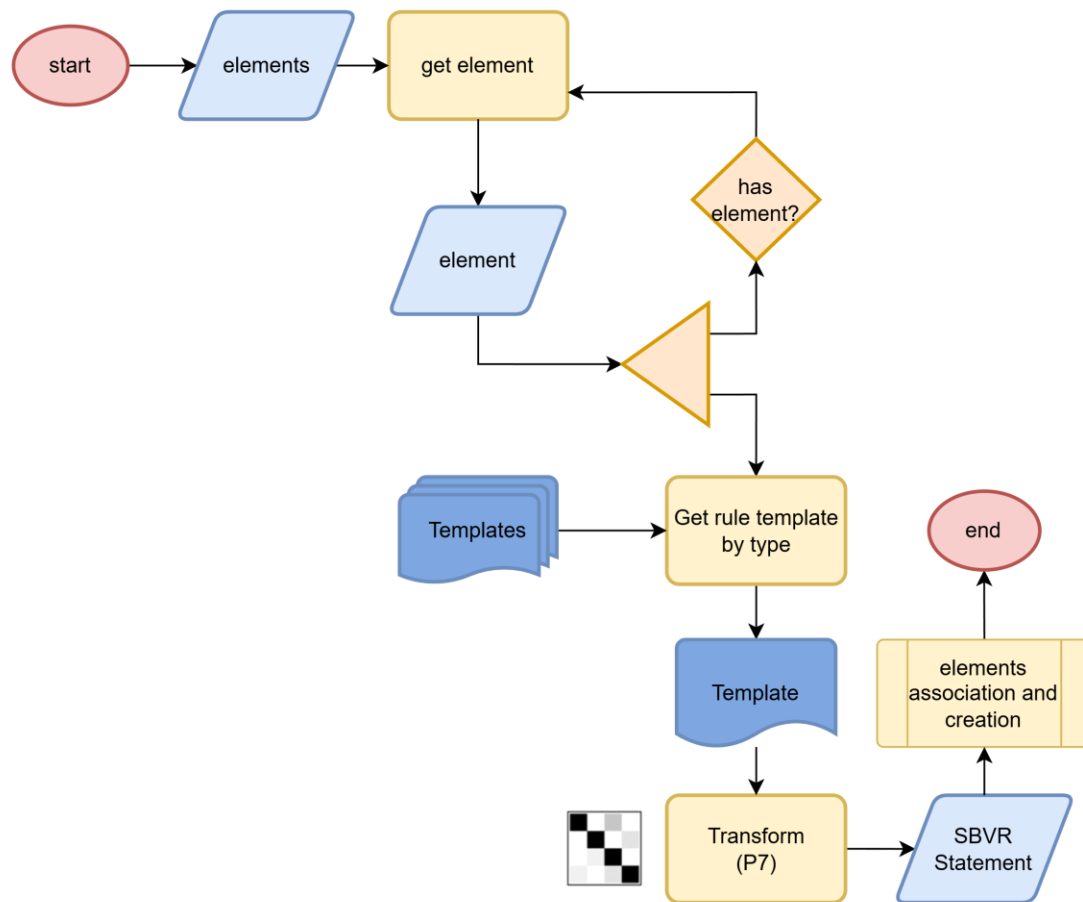
# Padrões para mapeamento de declarações SBVR

**TABLE 2.** Patterns to be identified and mapped to SBVR elements.

	Pattern	Example
<b>Pattern expressing explicit clauses</b>		
P1	Subject + 'to be' + Adjective	<p>Rental is open</p>
P2	Subject + 'to be' + Noun	<p>City branch is agency</p>
P3	Subject + Verb + Direct Object	<p>driver has license</p>
P4	Passive Nominal Subject + 'to be' + Verb in Past Participle + no agent	<p>rule is violated</p>
P5	Passive Nominal Subject + 'to be' + Verb in Past Participle + agent	<p>Rule is violated by rental</p>
<b>Pattern expressing implicit clauses</b>		
P6	Noun (NN) + Verb in past participle (VBN) as a modifier Implicit clauses: VBN + NN + "is" + NN & NN + "is" + VBN	<p>requested car</p>
P7	Noun (NN) + Adjective (JJ) as a modifier Implicit clauses: JJ + NN + 'is' + NN & NN + 'is' + JJ	<p>cheap car</p>



Figura 19 - Algoritmo “nlp2sbvr”.



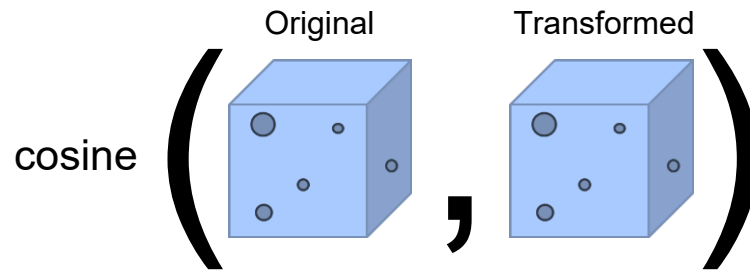
Fonte: Elaborado pelo autor



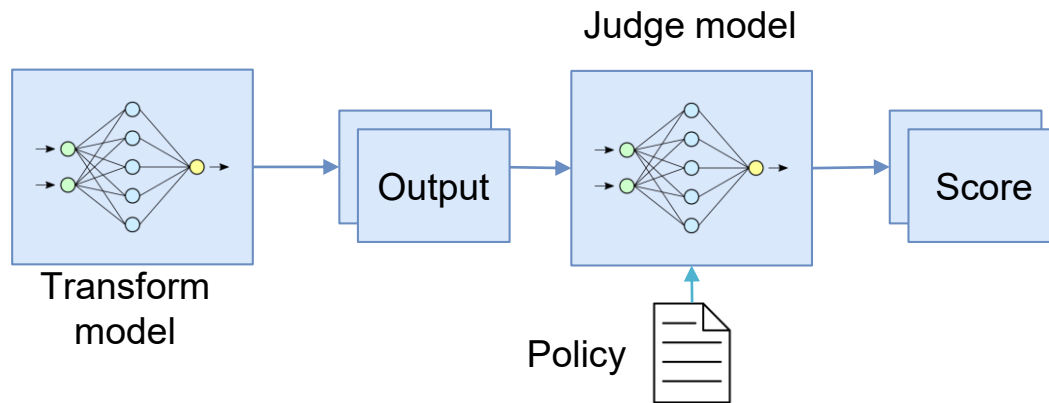
# Validação

	Actual	
	positive	negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Métricas estatísticas



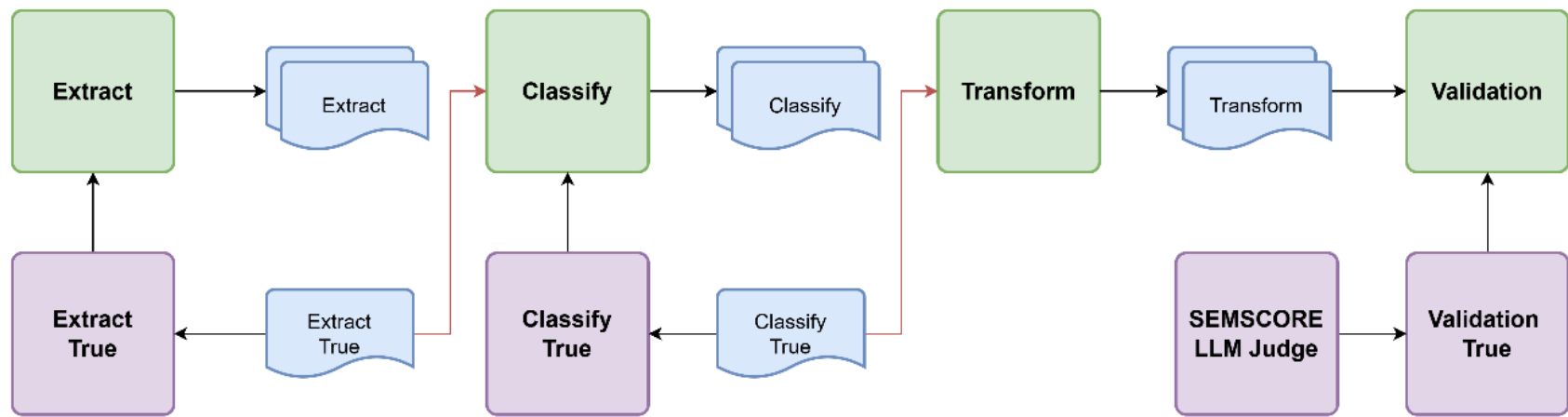
SEMSCORE



LLM as a Judge

# CFR2SBVR - Processos

Figura 23 - Processo CFR2SBVR

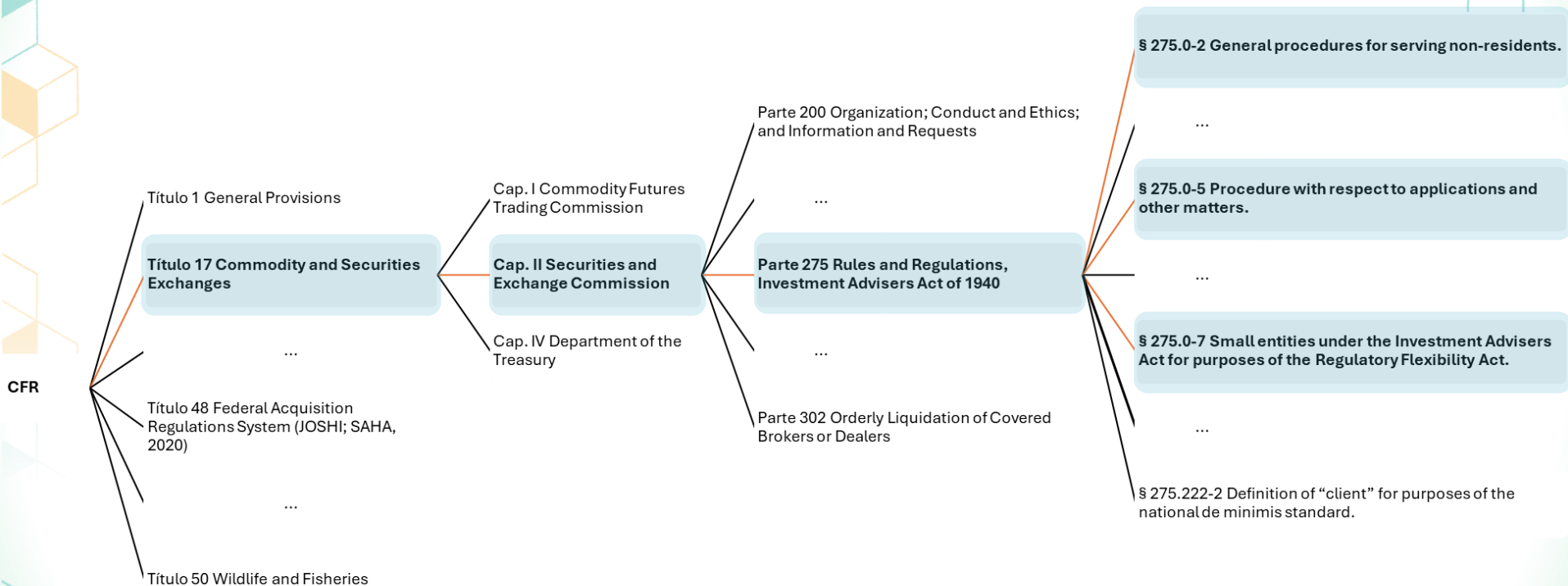


Fonte: Elaborado pelo autor

1.	Revisão da quali	5 min
2.	Experimento	15 min
3.	<b>Resultados</b>	8 min
4.	Discussão e conclusão	5 min
5.	Trabalhos futuros	2 min



# Conjunto de dados



Fonte: Elaborado pelo autor

# Conjunto de dados (Extração)

Tabela 2 - Conjunto de dados ouro para validação (P1)

<b>Id</b>	<b>Elementos</b>	<b>Tipo de fatos</b>	<b>Regras operativas</b>	<b>Símbolos verbais</b>	<b>Nomes</b>	<b>Termos</b>
§ 275.0-2	7	5	2	21	3	27
§ 275.0-5	5	1	4	12	3	20
§ 275.0-7	10	10	0	28	2	30
<b>Total</b>	<b>22</b>	<b>16</b>	<b>6</b>	<b>61</b>	<b>8</b>	<b>77</b>

Fonte: Elaborado pelo autor

Tabela 3 - Termos e nomes com definição

<b>Id</b>	<b>Termos totais</b>	<b>Termos com definição</b>	<b>Nomes totais</b>	<b>Nomes com definição</b>
§ 275.0-2	27	4	3	2
§ 275.0-5	20	16	3	3
§ 275.0-7	30	8	2	0
<b>Total</b>	<b>77</b>	<b>28</b>	<b>8</b>	<b>5</b>

Fonte: Elaborado pelo autor

Tabela 4 - Resumo das métricas de classificação.

<b>Elementos</b>	<b>Cont.</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b>Acurácia</b>	<b>F1</b>
Tipos de fatos subtipo	160	0,97	0,90	0,90	0,93
Nomes subtipo	50	0,85	0,76	0,76	0,76
Regras operativas tipo	60	0,94	0,93	0,93	0,93
Regras operativas subtipo	60	0,86	0,83	0,83	0,84
Termos subtipo	280	0,97	0,90	0,90	0,94

Fonte: Elaborado pelo autor.

Tabela 7 - Estatística das métricas combinadas.

Elementos	Cont.	SEMSCORE					LLM as a Judge				
		Média	DP	Min	75%	Max	Média	DP	Min	75%	Max
Tipos de fatos	160	0,87	0,05	0,71	0,92	0,97	0,92	0,06	0,70	0,95	0,95
Nomes	50	0,88	0,03	0,81	0,90	0,91	0,95	0,01	0,90	0,95	0,95
Regras operativas	60	0,91	0,02	0,87	0,92	0,93	0,90	0,05	0,75	0,95	0,95
Termos	280	0,85	0,08	0,50	0,90	0,96	0,92	0,05	0,60	0,95	1,00

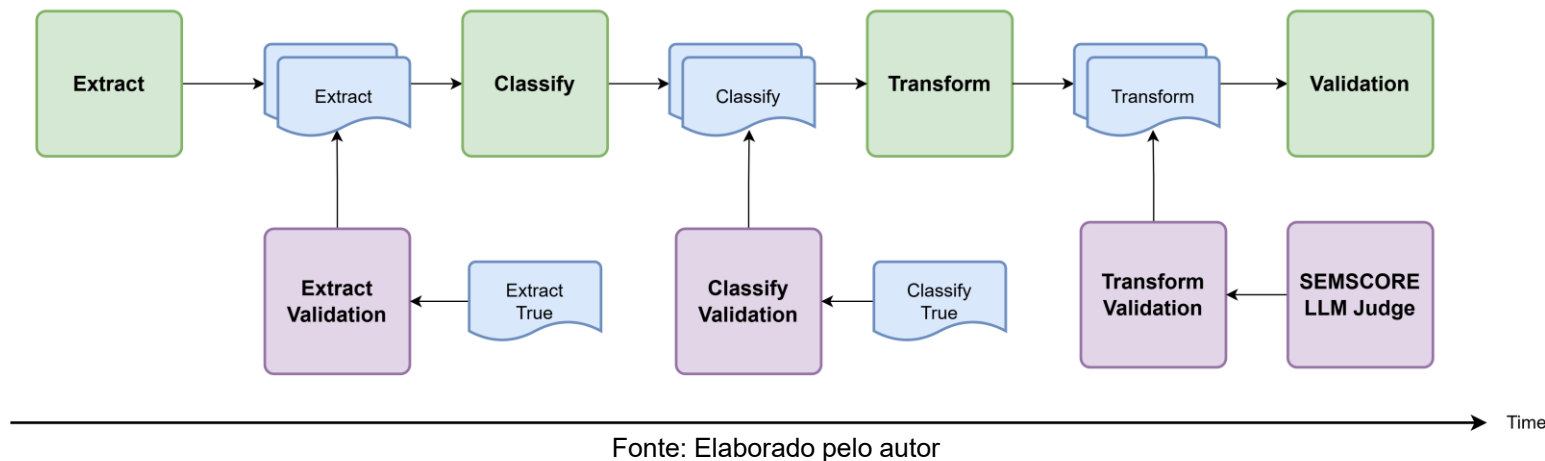
Fonte: Elaborado pelo autor.



Quadro 34 - Resultado dos autores.

<b>Autores</b>	<b>Precisão (regras def.)</b>	<b>Precisão (regras op.)</b>	<b>Conj. dados</b>	<b>Qde Extraídas / Avaliadas</b>	<b>Tempo de Execução</b>
Esta pesquisa	85%-95%	90%-95%	Título 17, Cap. II, Parte 275. CFR	770 termos, 80 nomes, 160 tipos de fatos, 610 verbos, 60 regras operativas.	~40 min
Chittimalli et al. (2020)	89%-97%	84%	Conjuntos de teste simulados (KYC e EU-Rent)	KYC: 548 entidades, 350 fatos, 638 regras. EU-Rent: 277 entidades, fatos 177, 64 regras	EURent: 3 min KYC: 10 min
Roychoudhury et al. (2017)	Não informado	Não informado	MiFID II e KYC	187 regras (KYC)	9h (KYC com SE)
Abi-Lahoud et al. (2013)	Não informado	Não informado	76 FR 45403 - 18 páginas	300 substantivos	Não especificado (totalmente manual)
Haj et al. (2021)	89%-96%	67%-95%	Extraído de Livros 1, 2, 3	152 sentenças (preferências, reservas e empréstimos)	Não especificado (automatizado)
Joshi e Saha (2020)	Não informado	Não informado	Título 48 CFR	9.084 expressões deonticas (710 permissões, 698 obrigações, 479 proibições, 149 dispensas)	Não informado

Fonte: Elaborado pelo autor



$$P(Sucesso) = \bar{P}_e \cdot \bar{P}_c \cdot \bar{P}_t = 0,73$$

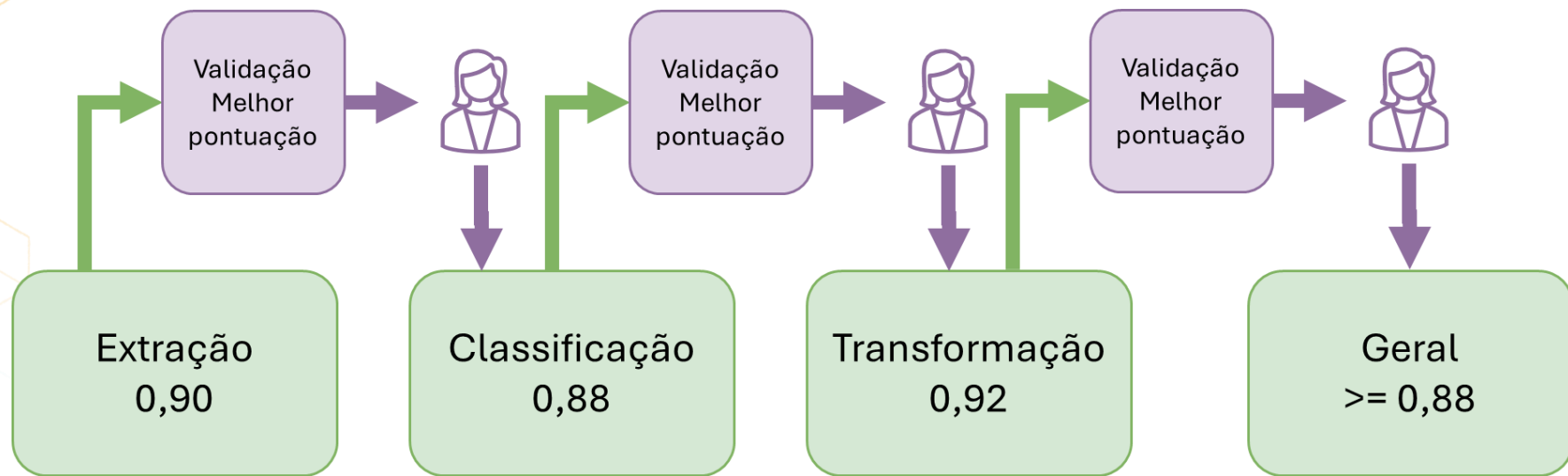
Onde:

$\bar{P}_e = 0,90$  precisão média da extração

$\bar{P}_c = 0,88$  precisão média da classificação

$\bar{P}_t = 0,92$  precisão média da transformação

Figura 38 - CFR2SBVR semi-automático



Fonte: Elaborado pelo autor

1.	Revisão da quali	5 min
2.	Experimento	15 min
3.	Resultados	8 min
4.	Discussão e conclusão	5 min
5.	Trabalhos futuros	2 min

✓ **Objetivo:** Propor, implementar e validar um método automatizado que empregue técnicas avançadas de NLP para transformar regulamentos Code of Federal Regulations (CFR), escritos em NL, para regras e vocabulário em inglês estruturado (SE) conforme a especificação SBVR, abreviada para CFR2SBVR.

✓ **Pergunta:** Como a aplicação de técnicas avançadas de processamento em linguagem natural e ontologias, como a FIBO, **podem aprimorar a transformação automatizada** de regulamentos financeiros em regras e vocabulário estruturados, conforme a especificação SBVR?

## Discussão e conclusão

**Hipótese:** A aplicação de um processo **automatizado** baseado em técnicas avançadas de NLP com um modelo de IA probabilístico pré-treinado, o uso de grafo de conhecimento para representar textos legais e o suporte da **FIBO** para o vocabulário deverão resultar em **uma melhora significativa** na precisão e eficiência da transformação.

### Automação

- Automação completa não é confiável, boas pontuações não são suficientes, recomendado o uso como ferramenta para SMEs.

### Ontologias

- Ontologias como FIBO e FRO não foram essenciais para a mecânica da transformação.
- LLMs têm potencial para transformar textos em regras formais sem entender o significado, apenas os papéis dos termos.
- O vocabulário da FIBO tem um desalinhamento com a CFR.

# Discussão e conclusão

## Ameaças a validade da pesquisa

- **Interna:** Dependência de um conjunto de dados padrão-ouro potencialmente enviesado, mitigada por revisão iterativa e comparação com métodos alternativos (métricas quantitativas, similaridade semântica e inspeção manual).
- **Externa:** Limitação do escopo a três seções da CFR dificulta a generalização, não especificidade do algoritmos com a CFR e técnicas de adaptação de domínio.
- **Construto:** Risco de métricas quantitativas não capturarem totalmente a semântica, mitigado pela combinação de análises quantitativas e qualitativas.
- **Confiabilidade:** Variabilidade de modelos estocásticos foi mitigada com múltiplas execuções e disponibilização pública do código e dados para replicação.



# Discussão e conclusão

## Experimento CFR2SBVR

- Permitiu evidenciar uma viabilidade técnica e economicamente, com pontuações de precisão, acurácia, sensibilidade e F1 score **acima de 0,85** em média.
- Custo por execução (3 seções) de aproximadamente **USD 25.00**, com tempo médio inferior a **4 minutos**.
- Sem adaptações, a precisão geral poderia cair para cerca de 0,73, mas a incorporação do processo de validação (IA) e a avaliações de especialistas pode-se elevar essas pontuações.

## Contribuição acadêmicas

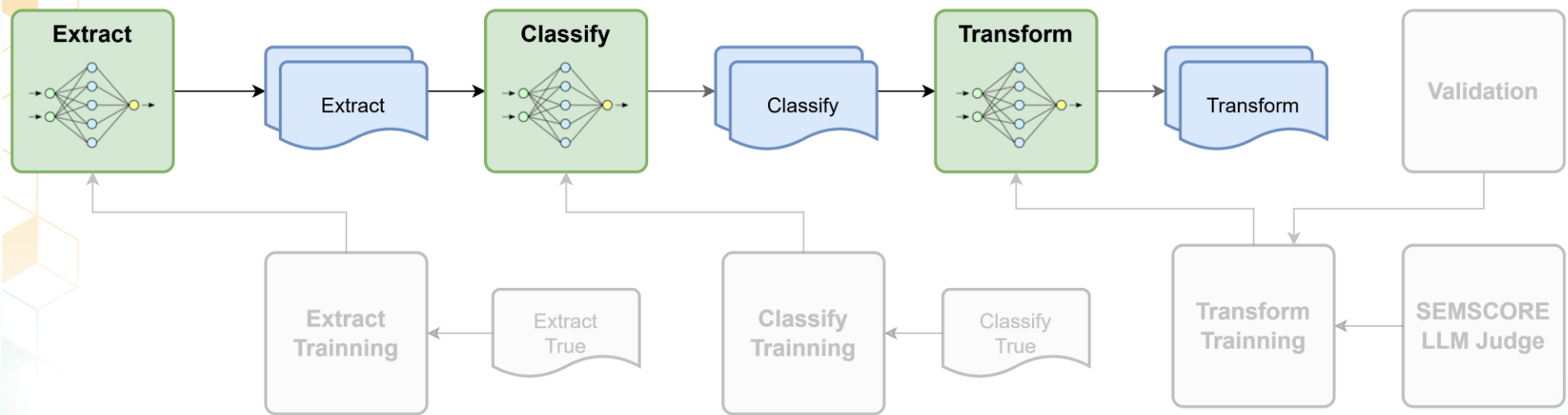
1. Aplicação do SBVR no setor financeiro e em processos de GRC.
2. A combinação de técnicas e estratégias para a transformação de NL para SBVR.
3. CFR2SBVR como ferramenta de suporte para SMEs.
4. Código-fonte e conjunto de dados para futuras comparações e reprodutibilidade.

1.	Revisão da quali	5 min
2.	Experimento	15 min
3.	Resultados	8 min
4.	Discussão e conclusão	5 min
5.	Trabalhos futuros	2 min



# Ajuste fino dos modelos

Trabalhos futuros



Fonte: Elaborado pelo autor.



# Ajuste fino dos modelos

Trabalhos futuros

**Reinforcement  
Learning with  
Human Feedback  
(RLHF)**

**Reinforcement  
Learning from AI  
Feedback (RLAIF)**

**Direct Preference  
Optimization  
(DPO)**

**Kahneman-  
Tversky  
Optimization  
(KTO)**

Fonte: Elaborado pelo autor.

# Trabalhos futuros

## GRC

- Ampliar o escopo do trabalho para as demais atividades de GRC;
- Pesquisa estruturada com fornecedores de soluções para GRC e o uso de IA

## Taxonomias e ontologias

- Melhorias na taxonômica de Witt (2012) e a incorporação de posições jurídicas de Hohfeld (1917).
- Desenvolvimento de ontologia SBVR. Mapeamento entre SBVR e MVF para dicionários terminológicos, ampliando a aplicabilidade do modelo.

## Arquitetura

- Outros LLMs, modelos com raciocínio embutido, código-aberto, fechado, com mais / menos parâmetros
- Uso de sistemas multiagentes
- Ajuste fino dos LLMs por processo

## Processos

- Implementação do processo de validação no CFR2SBVR.
- Interfaces gráficas para interação entre SMEs e LLMs e registro de feedbacks.

## Integração

- Sinergia de abordagens como *Domain-Driven Design* (DDD) e SBVR, como as definições estruturadas do SBVR podem contribuir com as definições de domínios e contextos do DDD..

## **Aplicação de demonstração**

<https://cfr2sbvr.streamlit.app/>

## **Repositório de código**

<https://github.com/asantos2000/master-degree-santos-anderson>

