

O Uso da Inteligência Artificial no Schema Matching: Um Mapeamento Sistemático

Ricardo H. D. Borges¹, Leonardo Andrade Ribeiro¹, Valdemar V. Graciano Neto¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, s/n - Chácaras Califórnia, Goiânia - GO - Brazil

{ricardoborges, laribeiro, valdemarneto}@ufg.br

Abstract. *With the increasing complexity of data integration due to the growth in both quantity and diversity of data sources, Schema Matching plays a pivotal role. In this challenging landscape, Artificial Intelligence (AI) emerges as a promising solution to enhance the efficiency of Schema Matching. This article presents the results of a systematic literature review, investigating the most widely used AI techniques and algorithms in Schema Matching applications. The insights gained offer valuable guidance for researchers and professionals seeking to improve data integration through Schema Matching.*

Resumo. *Com a crescente complexidade da integração de dados devido ao aumento em sua quantidade e diversidade, o Schema Matching desempenha um papel fundamental. Nesse cenário desafiador, a Inteligência Artificial (IA) surge como uma solução promissora para aprimorar a eficiência do Schema Matching. Este artigo apresenta os resultados de um mapeamento sistemático da literatura, investigando as técnicas e algoritmos de IA mais utilizados em aplicações de Schema Matching. Os insights obtidos oferecem orientação valiosa para pesquisadores e profissionais que buscam aprimorar a integração de dados por meio do Schema Matching.*

1. Introdução

A integração de dados é crucial para muitas organizações que precisam combinar dados de diferentes fontes para transformar em informações. No entanto, a integração de dados pode ser um desafio especialmente quando os dados provêm de fontes com esquemas diferentes. Realizada de forma desordenada, a integração de dados pode levar a erros e inconsistência. Nesse contexto, o *Schema matching* se apresenta como uma abordagem essencial na integração de dados que consiste na identificação de correspondências entre esquemas de diferentes fontes de dados [Rahm and Bernstein 2001]. Esse processo consiste na identificação e estabelecimento de correspondências entre os elementos de dois ou mais esquemas de dados diferentes. Essa correspondência permite a integração e o compartilhamento de informações entre sistemas ou fontes de dados heterogêneas, podendo ser aplicado de forma automática, o que leva a uma integração mais eficiente e precisa; e manual. [Bilke and Naumann 2005].

A contribuição da Inteligência Artificial (IA) para aprimorar o *Schema matching* é cada vez mais proeminente. Por meio de técnicas de IA, é possível automatizar grande parte do processo de correspondência de esquemas (*Schema Matching*), reduzindo a dependência de intervenção manual intensiva. Isso não apenas acelera a integração de dados,

mas também aprimora a precisão, identificando correspondências sutis que poderiam passar despercebidas de outra forma. A IA também tem a capacidade de aprender com as decisões de mapeamentos anteriores, refinando continuamente suas estratégias e adaptações à medida que novos dados e desafios surgem.

No entanto, um mapeamento sistemático sobre a aplicação da IA no *Schema matching* é fundamental para compreender e sintetizar de maneira completa o estado atual do conhecimento nessa convergência. Dado o constante desenvolvimento das técnicas de IA, tal mapeamento favorecerá uma visão global das abordagens, metodologias e tendências específicas desse domínio, auxiliando na identificação de lacunas de pesquisa, pontos fortes e limitações.

A principal contribuição deste artigo é a apresentação dos resultados de uma abordagem sistemática para mapear as técnicas de IA aplicadas ao *Schema Matching*. No total, 644 estudos foram inicialmente identificados pela *string* de busca, dos quais 68 foram incluídos com base em critérios específicos de seleção e *snowballing*. Notavelmente, as subáreas de *Deep Learning* (DL) e *Natural Language Processing* (NLP) emergiram como as mais amplamente utilizadas na literatura investigada. Além disso, outras técnicas não diretamente relacionadas à IA também foram aplicadas em conjunto com as técnicas de IA, evidenciando a interdisciplinaridade das abordagens utilizadas.

Este artigo está assim organizado: a Seção 2 detalha os métodos adotados no planejamento e execução do protocolo do Mapeamento Sistemático da Literatura (MSL); a Seção 3 reporta resultados obtidos ao longo da condução do MSL e a análise desses resultados; a Seção 5 descreve as ameaças à validade do MSL e, por fim, a Seção 6 apresenta as considerações finais e trabalhos futuros.

2. Metodologia

Para desenvolver esta pesquisa, foi utilizado o Mapeamento Sistemático da Literatura (MSL) de acordo com o protocolo proposto por Fabbri et al [Fabbri et al. 2013] e [Petersen et al. 2015]. O protocolo é composto por três etapas principais: Planejamento, Condução e Publicação dos Resultados, apresentado na Figura 1.

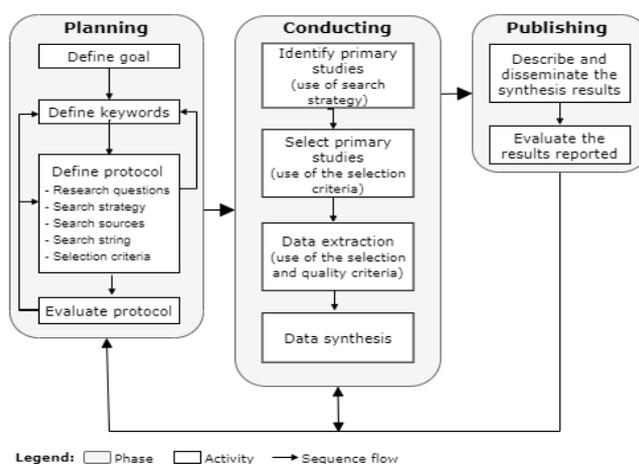


Figura 1. Fases e atividades do MSL, adaptado de Fabbri et al. [Fabbri et al. 2013].

2.1. Questões de Pesquisa

As questões de pesquisa que expressam os objetivos deste mapeamento foi formulada seguindo os critérios especificados por PICOC (Population, Intervention, Comparison, Outcomes, Context) definido em [Budgen and Brereton 2006]. Por tratar-se de um mapeamento, apenas PIO foi utilizado e a Tabela 1 exibe os detalhes.

Tabela 1. Critérios do PIO

População	Schema Matching, Inteligência Artificial, Machine Learning, Deep Learning, Natural Language Processing
Intervenção	métodos/técnicas/tecnologias/ferramentas/padrões
Resultados	Técnicas ou algoritmos utilizados

No âmbito da Inteligência Artificial, foram exploradas três questões de pesquisa que abordam cada uma das suas subáreas: Deep Learning (DL), Machine Learning (ML) e Natural Language Processing (NLP). Além disso, uma quarta questão de pesquisa foi examinada em relação a técnicas que não estão estritamente vinculadas a nenhuma subárea específica.

É possível observar que DL está situado dentro do domínio do ML, uma vez que DL é uma abordagem específica que se baseia em redes neurais profundas para o aprendizado de máquina. No entanto, o NLP é representado como uma área separada, pois é um campo especializado que se concentra exclusivamente no processamento e compreensão da linguagem natural. Embora o NLP utilize técnicas de Inteligência Artificial para seu funcionamento, é importante reconhecer sua singularidade devido ao foco específico na linguagem e em desafios linguísticos complexos. A seguir as questões de pesquisa (QP) levantadas durante esta fase e sua justificativa.

QP1: *Quais técnicas específicas de Deep Learning são utilizadas na aplicação de Schema Matching?*

- Procura investigar quais técnicas exclusivamente de Deep Learning estão sendo aplicadas em *Schema Matching*.

QP2: *Quais técnicas específicas¹ de Machine Learning têm sido utilizadas na aplicação de Schema Matching?*

- Procura investigar quais técnicas exclusivamente de Machine Learning estão sendo aplicadas em *Schema Matching*.

QP3: *Quais técnicas específicas² de NLP têm sido utilizadas na aplicação de Schema Matching?*

- Procura investigar quais técnicas exclusivamente de NLP estão sendo aplicadas em *Schema Matching*.

QP4: *Quais técnicas de IA têm sido utilizadas na aplicação de Schema Matching?*

- Procura investigar quais técnicas exclusivamente de IA que não se enquadram em DL, ML e NLP estão sendo aplicadas em *Schema Matching*.

¹Existem técnicas de ML que não são de DL

²Existem técnicas de NLP que não são de ML e DL

2.2. Identificação dos Estudos

O mapeamento foi conduzido utilizando uma estratégia de busca automática, nas bases Scopus³, IEEEXplore⁴, ACM⁵ e Engineering Village⁶. Além disso, foi aplicada a técnica do *Snowballing* para os estudos selecionados.

String de Busca

Fase fundamental para incluir os termos que são pertinentes às questões de pesquisa, ou seja, aqueles relacionados às palavras-chave: inteligência artificial, deep learning, machine learning e *Schema Matching*. Além disso, foi incluído de sinônimos para ampliar a abrangência dos resultados.

A *string* de busca definida para este trabalho foi:

(“deep learning”) OR (“Machine Learning”) OR (“Artificial Intelligence”) AND
 (“schema matching”) OR (“ontology matching”) OR (“ontology alignment”)

Durante a aplicação da *String* de busca, a mesma foi aprimorada continuamente, realizando ajustes e refinamentos com base nos resultados obtidos e na análise dos estudos encontrados. Ao longo desse processo de calibração da *String*, foi possível observar a evolução da *String* de busca, à medida que se faz o refinamento dos termos-chave, considerando sinônimos, variações linguísticas e acrônimos relevantes para a área de estudo. Como mostrado a seguir:

- Versão 1:** (“deep learning”) AND ((“schema matching”) OR (“ontology matching”) OR (“entity matching”));
- Versão 2:** (“deep learning”) AND ((“schema matching”) OR (“ontology matching”));
- Versão 3:** ((“deep learning”) OR (“Deep Neural Network”)) AND ((“schema matching”) OR (“ontology matching”));
- Versão 4:** ((“deep learning”) OR (“Machine Learning”) OR (“Artificial Intelligence”)) AND ((“schema matching”) OR (“ontology matching”) OR (“ontology alignment”)).

2.3. Critérios de Seleção

A etapa de definição dos critérios de seleção incluem os critérios de inclusão e exclusão, que são estabelecidos para orientar a seleção dos estudos que serão lidos na íntegra. A seguir, apresentamos os critérios de inclusão (CI) e critérios de exclusão (CE) utilizados:

- CI1:** O estudo apresenta algoritmos ou técnicas de Deep Learning aplicado a Schema Matching/Ontology Matching/Ontology Alignment;
- CI2:** O estudo apresenta algoritmos ou técnicas de Machine Learning aplicado a Schema Matching/Ontology Matching/Ontology Alignment;
- CI3:** O estudo apresenta algoritmos ou técnicas de NLP aplicado a Schema Matching/Ontology Matching/Ontology Alignment;
- CI4:** O estudo apresenta conjunto de algoritmos ou técnicas IA aplicado a Schema Matching/Ontology Matching/Ontology Alignment.

CE1: O estudo não é um estudo primário;

³<https://www.scopus.com>

⁴<https://ieeexplore.ieee.org>

⁵<https://www.acm.org>

⁶<https://www.engineeringvillage.com>

- CE2:** O estudo não está disponível para acesso gratuito;
- CE3:** O estudo não está escrito em Inglês ou Português;
- CE4:** O estudo não foi publicado nos últimos 5 anos;
- CE5:** O estudo não apresenta algoritmos ou técnicas de Inteligência Artificial aplicado a Schema Matching/Ontology Matching/Ontology Alignment;
- CE6:** O estudo não é um artigo, ou seja, não foi revisado por partes, sendo uma literatura cinza.

Estudos que atenderam pelo menos um critério de inclusão foram incluídos na seleção inicial e os estudos que atendem a pelo menos um critério de exclusão são excluídos da seleção inicial. Inicialmente, foi aplicado o critério de exclusão de limitar os estudos aos últimos 10 anos. No entanto percebeu-se que o tema em estudo tem passado por uma rápida evolução especialmente a partir dos anos de 2018 e 2019, com o surgimento de estudos que aplicam diversas técnicas inovadoras. Diante dessa percepção, foi reduzido o escopo do mapeamento para os últimos 5 anos.

A partir da extração inicial de 644 estudos e importado na ferramenta⁷, foram eliminados os duplicados (151 estudos). Assim, realizando a leitura de todos os títulos, resumos e até conclusões de cada estudo, a fim de identificar os estudos que realmente atenderam o tema proposto. Após isso, foram extraídos a seleção inicial dos estudos para realizar a leitura completa. Nesta seleção foram extraídos inicialmente 82 estudos. Porém, com a leitura completa percebeu-se que ainda havia estudos que fugiam do tema proposto. Desta forma, os estudos não relevantes foram excluídos, restando ao final 59 estudos para extração.

2.4. Extração dos Dados

Durante esta etapa, interrompeu-se o uso da ferramenta, alterando o foco na manipulação desta extração de dados através de planilhas⁸, com o propósito de obter uma perspectiva sobre a evolução das seleções ao longo do protocolo. Essas planilhas foram criadas a partir da plataforma Parsif.al. Durante o processo de extração, todas as questões de pesquisa foram abordadas, e os estudos foram categorizados com base em suas respostas para cada uma delas.

2.5. Snowballing

Com os estudos identificados no MSL, foi aplicado o método *Snowballing*. Sendo examinadas as referências bibliográficas dos estudos identificados e também por meio da busca de estudos que os citam, assim, sendo possível identificar 26 novos estudos relevantes que não haviam sido incluídos no MSL. Esses 26 estudos foram submetidos ao mesmo protocolo de análise do MSL e, ao final, apenas 9 foram considerados aptos para o trabalho em questão, totalizando **68 estudos**. A aplicação do método Snowball permitiu a ampliação da base de estudos e a identificação de novas fontes relevantes para a pesquisa.

3. Resultados

Os estudos relevantes estão organizados nesse *link*⁹, onde é possível visualizar a referência completa do estudo e um código de identificação para os estudos acrescido de um valor

⁷Parsif.al

⁸Link das planilhas

⁹<https://github.com/articlesgo/IA-SM/blob/main/estudos.pdf>

numérico, que será utilizado como referência. Os estudos foram disponibilizados externamente por questões de restrição de espaço.

3.1. Quais técnicas ou algoritmos de Deep Learning têm sido utilizados na aplicação de *Schema Matching*?

Entre as técnicas ou algoritmos da Deep Learning encontradas destacam-se as Redes Neurais Siameses (SNN) com 23,3%, Redes de Memória de Curto Prazo Longa (LSTM) com 26,7% e Redes Neurais Convolucionais (CNN) com 6,7%. Além disso, o Multi-layer Perceptron (MLP) com 13,3%, Rede Neural Recorrente (RNN) e Gráfico de Redes Convolucionais (GCN) com 10,0%. Outros como a Rede Neural Recursiva (RvNNs), Competitive Learning e Multi-Input com 3,3%. Veja na Figura 2.

As proporções indicam que as LSTM e as SNNs são as técnicas de Deep Learning mais prevalentes no contexto de *Schema matching*, seguidas por MLPs, RNNs, GCNs e CNNs. Essas informações sugerem quais técnicas são mais frequentemente empregadas ou consideradas úteis para o *Schema Matching*.

Tabela 2. Estudos selecionados

SNN	CNN	LSTM	MLP	RNN	GCN	RvNNs, Competitive Learning e MIMO
E01, E02, E03, E04, E05, E06 e E07	E08 e E09	E10, E11, E12, E13, E14, E15 e E16	E17, E18, E19 e E08	E12, E13 e E16	E09, E20 e E21	E14, E07 e E22

3.2. Quais técnicas ou algoritmos de Machine Learning têm sido utilizados na aplicação de *Schema Matching*?

A Figura 3 apresenta uma variedade de técnicas e algoritmos de Machine Learning identificados. Entre as técnicas mais comuns, destacam-se o classificador Naive Bayes, com 20,7%, a árvore de decisão (Decision Tree) com 17,2% e a floresta aleatória (Random Forest), com 13,8%. Além disso, como k-means com 13,8%, JRip, SVM com 3,4%, 6,9% respectivamente. Regressão com 10,3%, Adaboost com 6,9% e K-Nearest com 3,4%. Os resultados desses estudos também evidenciam a eficácia do uso de técnicas e algoritmos de Machine Learning em aplicações de *Schema Matching*, o que pode ajudar a melhorar a qualidade e a precisão de sistemas que lidam com dados heterogêneos.

Tabela 3. Estudos selecionados

Naive Bayes	Decision Tree	Random Forest	k-means	JRip, SVM e C4.5	Adaboost	K-Nearest	Regression, Gaussian mixture e Algoritmo evolutivo
E23, E24, E25, E26, E27, E28 e E29	E23, E30, E31 e E32	E33, E23, E34 e E35	E36, E28, E37 e E38	E26 e E31	E31 e E23	E23 e E31	E33, E23, E24, E67 e E68

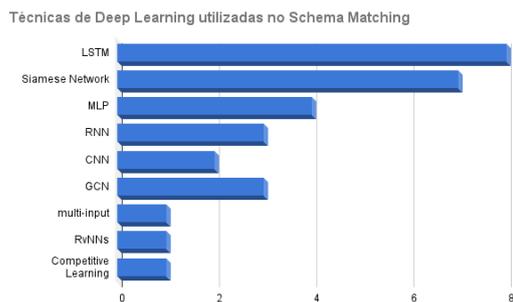


Figura 2. Técnicas de Deep Learning têm sido utilizados no Schema Matching

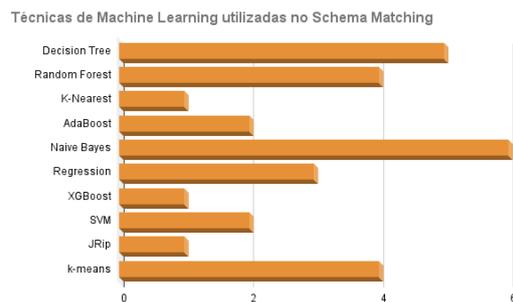


Figura 3. Técnicas de Machine Learning utilizadas no Schema Matching

3.2.1. Quais técnicas ou algoritmos de NLP têm sido utilizados na aplicação de Schema Matching?

A Figura 5 relata diversas técnicas e algoritmos de Processamento de Linguagem Natural (NLP). Dentre as técnicas mais comuns encontradas, destacam-se o BERT (Bidirectional Encoder Representations from Transformers) um modelo de linguagem pré-treinado baseado na arquitetura Transformers, com 17,6% dos estudos. Além do BERT, outros algoritmos de NLP também foram identificados. Por exemplo, o Word2Vec com 41,2%, enquanto o FastText com 11,8%, o GloVe com 11,8%. TransE e StarSpace com 5,9%. Byte-Pair Encoding (BPE) e abordando também o uso de análise de sentimentos, com 2,9% e também técnicas de embedding sem especificar o algoritmo.

Portanto, o Word2Vec é a técnica mais amplamente utilizada, seguida pelo BERT, FastText e GloVe, com outras técnicas também sendo relevantes, embora com menor frequência de uso.

Tabela 4. Estudos selecionados

BERT	Word2Vec	FastText	GloVe	TransE	StarSpace	BPE	Embeddings (Sem especificar)
E11, E39, E40, E41, E42 e E43	E44, E45, E46, E47, E34, E48, E42, E49, E50, E21, E51, E16, E52 e E53	E40, E54, E42 e E51	E55, E42, E51 e E52	E21, E56	E36 e E57	E58	E01, E02, E06 e E59

3.2.2. Quais técnicas ou algoritmos de Inteligência artificial têm sido utilizados na aplicação de Schema Matching?

A Figura 4 mostra que foi encontrado também estudos isolados, onde utilizaram algoritmos mais generalistas da própria Inteligência Artificial. Em alguns desses estudos, não foi especificado qual algoritmo foi utilizado, enquanto outros mencionaram apenas a subárea (Machine Learning ou Deep Learning) sem detalhar o algoritmo específico, com 71,4%, Machine Learning, 14,3% Deep Learning e o Artificial Bee Colonies com 14,3%.

Tabela 5. Estudos selecionados

IA Sem especificar o algoritmo	Deep Learning	Grasshopper Algorithm	Bee Colonies
E60, E61, E62, E63 e E64	E51	E65	E66

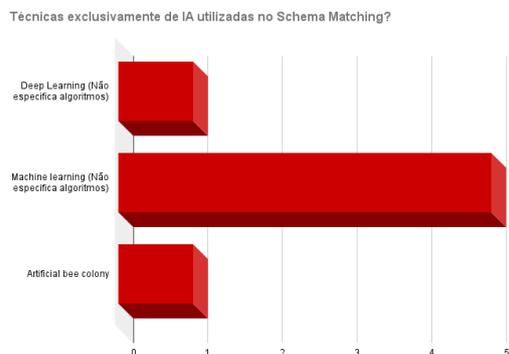


Figura 4. Técnicas de IA utilizadas no *Schema Matching*

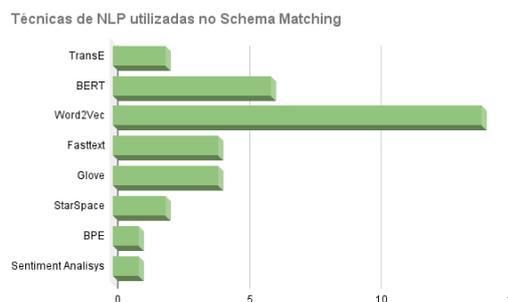


Figura 5. Técnicas de NLP utilizadas no *Schema Matching*

4. Discussão

Durante a análise dos estudos, observou-se a importância e o amplo uso de IA aplicada no *Schema Matching* para a integração de dados. A complexidade e o custo de identificar correspondências em uma grande massa de dados pode expandir significativamente o impacto de falhas, tornando o custo e o tempo necessários para resolvê-lo um problema relevante, principalmente quando executada de forma manual.

Além disso, com os dados extraídos do mapeamento foi constatado a ampla variedade de técnicas de Inteligência Artificial (IA) aplicadas no *Schema Matching*, sendo uma grande tendência o uso de técnicas de DL e NLP, como na Figura 6. Além disso, também observou-se a presença de abordagens como similaridade e análise de ontologias semânticas junto com técnicas de IA e o uso de mais de uma técnica de IA durante o processo de *Schema Matching*. Essa diversidade de técnicas demonstra que ainda não existe uma única solução ideal para o problema de *Schema Matching*. Pelo contrário, é possível aplicar uma combinação de várias técnicas, adaptando-as ao contexto específico e às necessidades do projeto em questão. A escolha da abordagem mais adequada dependerá das características dos dados, dos requisitos do problema e das metas de qualidade estabelecidas.

Uma das principais lacunas reside na capacidade de compreender semântica e contexto complexos, uma vez que o *Schema Matching* muitas vezes vai além da mera correspondência de palavras. A escalabilidade e o desempenho também são preocupações, especialmente com o aumento na quantidade de dados e esquemas a serem correspondidos. Além disso, manter a correspondência atualizada e automatizada é uma tarefa desafiadora. Isso inclui a necessidade de aprender com dados limitados e medir o desempenho de maneira objetiva. A diversidade na estrutura e formato dos esquemas, juntamente com a interoperabilidade em ambientes heterogêneos, é outra área que precisa de atenção. Além disso, sistemas que possam aprender com o *feedback* dos usuários e melhorar continuamente a correspondência de esquemas são necessários.

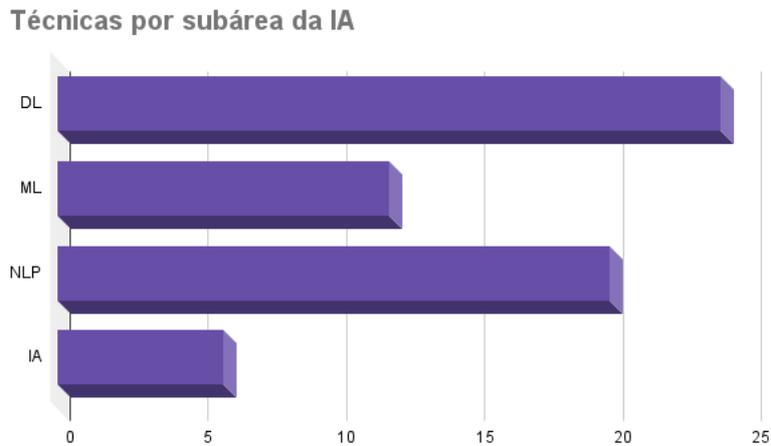


Figura 6. Técnicas por subárea da IA utilizadas no *Schema Matching*

No entanto, apesar do amplo uso das subáreas de DL e NLP na pesquisa de *Schema Matching*, essas lacunas servem como indicações claras das áreas onde a pesquisa neste campo ainda tem muito a explorar e aprimorar.

5. Ameaças à Validade

Ameaças à validade deste estudo foram identificadas e foram categorizados de acordo com [Hyman 1982] e [Wohlin et al. 2012]:

Validade da construção: Pode haver uma possível exclusão de estudos relevantes. Para mitigar esse problema, foi realizada revisões de cada etapa de condução e extração de dados deste estudo mais de uma vez. A falta da especificidade do tema também foi uma ameaça encontrada, assim, encontramos estudos com foco em técnicas e algoritmos de IA aplicado ao *Schema Matching*, porém sem especificar o algoritmo utilizado. Sendo assim, conseguimos transpor estes resultados identificando pontos aplicados ao *Schema Matching*.

Validade Interna: Podem ter surgido devido aos métodos de busca escolhidos. Por exemplo, a opção de não incluir algumas bibliotecas digitais pode levar à exclusão de estudos relevantes e ao número relativamente baixo de estudos incluídos. Para mitigar essa ameaça, o protocolo foi previamente avaliado pelos orientadores para identificar possíveis erros.

Validade Externa: Questões externas como a indisponibilidade de estudos foram resolvidas por meio de pesquisa utilizando a Portal de Periódicos da CAPES, também utilizou-se a literatura cinza como Google Acadêmico, Google, Researchgate.

Validade de Conclusão: Possíveis ameaças estão relacionadas ao viés durante a condução e extração de dados, o que pode causar imprecisão na extração de dados, ameaçando a conclusão dos resultados do estudo. Para mitigar essas ameaças, foi apresentado os resultados em uma disciplina de Metodologia Científica e a um grupos de estudo e pesquisa.

6. Considerações Finais

Neste artigo, foram apresentados os resultados de um mapeamento sistemático sobre a aplicação da Inteligência Artificial (IA) no processo de *Schema Matching*. Foram explorados diversos estudos e pesquisas que utilizaram técnicas de IA em Deep Learning, Machine Learning e NLP, e identificado o uso de técnicas como Redes Neurais Convolucionais (CNN), Multi-layer Perceptron (MLP), dentre outros para abordar o desafio de *Schema Matching*.

Durante a discussão dos resultados, foi possível observar o amplo uso de IA no contexto do *Schema Matching*. Ficando claro que a IA pode ter um papel crucial na otimização do *Schema Matching*, permitindo a identificação de correspondências entre elementos de esquemas heterogêneos. Essas descobertas enfatizam a necessidade de continuar explorando e aprimorando as técnicas de IA aplicadas ao *Schema Matching*, com o objetivo de compará-las com abordagens não baseadas em IA. Esse enfoque contínuo na evolução das técnicas de IA certamente contribuirá para um avanço significativo no campo do *Schema Matching* e suas aplicações práticas.

Para futuras pesquisas é importante explorar o desenvolvimento de novos modelos de IA personalizados para abordar os desafios do *Schema Matching*, como o uso de abordagens híbridas que combinam técnicas de IA com métodos tradicionais de *Schema Matching*, onde podem oferecer um potencial significativo. A garantia de segurança e privacidade durante o processo de *Schema Matching* para dados críticos também são desafios adicionais que merecem atenção.

Referências

- Bilke, A. and Naumann, F. (2005). Schema matching using duplicates. pages 69–80.
- Budgen, D. and Brereton, P. (2006). Performing systematic literature reviews in software engineering. In *Proceedings of the 28th international conference on Software engineering*, pages 1051–1052.
- Fabbri, S. C. P. F., Felizardo, K. R., Ferrari, F. C., Hernandez, E. C. M., Octaviano, F. R., Nakagawa, E. Y., and Maldonado, J. C. (2013). Externalising tacit knowledge of the systematic review process. *IET Softw.*, 7(6):298–307.
- Hyman, R. (1982). Quasi-experimentation: Design and analysis issues for field settings (book). *Journal of Personality Assessment*, 46(1):96–97.
- Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology*, 64:1–18.
- Rahm, E. and Bernstein, P. (2001). A survey of approaches to automatic schema matching. *VLDB J.*, 10:334–350.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.