26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

XXVI ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - XXVI ENANCIB

ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento

REPRESENTAÇÃO TEMÁTICA COMPARADA EM GOVERNO DIGITAL

COMPARED THEMATIC REPRESENTATION IN DIGITAL GOVERNMENT

Fabiene Letízia Alves Furtado – Universidade Federal de Minas Gerais (UFMG)
Célia da Consolação Dias – Universidade Federal de Minas Gerais (UFMG)
Rodrigo Diniz Lara – Universidade Federal de Minas Gerais (UFMG)
Oseas Teixeira – Universidade Federal de Minas Gerais (UFMG)
Elisângela Cristina Aganette – Universidade Federal de Minas Gerais (UFMG)

Modalidade: Trabalho Completo

Resumo: A transformação digital dos serviços públicos amplia os desafios da organização e da recuperação da informação, especialmente diante da emergência de novos conceitos e da diversidade de objetos informacionais. Este estudo analisou comparativamente a representação temática de objetos textuais e não textuais do domínio do Governo Digital, considerando as contribuições da linguagem natural, do vocabulário controlado e da inteligência artificial quanto à aderência semântica, à especificidade terminológica e à ocorrência de inferências temáticas inadequadas. A pesquisa, de caráter aplicado e exploratório, adotou abordagem qualitativa e quantitativa e foi estruturada como um experimento de representação temática comparada. O corpus compreendeu 16 objetos, sendo 8 textuais e 8 não textuais, dos quais foram analisadas 174 ocorrências de termos. Os procedimentos envolveram a identificação de assuntos em linguagem natural, a validação terminológica e semântica com apoio de glossários especializados, a comparação com o Vocabulário Controlado do Governo Eletrônico e a análise dos termos gerados por uma ferramenta de inteligência artificial. Os resultados indicaram maior aderência entre a linguagem natural e os glossários especializados do que entre a linguagem natural e o VCGE, limitações deste na representação de conceitos emergentes e potencial da inteligência artificial como recurso de apoio, embora tenham sido identificadas divergências e inferências não verificadas. Conclui-se que, no domínio do Governo Digital, a articulação entre linguagem natural, vocabulário controlado e inteligência artificial, aliada à atualização terminológica e à validação humana, pode contribuir para uma representação temática mais consistente.
Palavras-chave: representação da informação; sistemas de organização do conhecimento; controle de vocabulário; governo eletrônico; inteligência artificial.

Abstract: The digital transformation of public services amplifies the challenges of organizing and retrieving information, especially given the emergence of new concepts and the diversity of information objects. This study comparatively analyzed the subject representation of textual and non-textual information objects in the Digital Government domain, considering the contributions of natural language, controlled vocabulary, and artificial intelligence regarding semantic adherence, terminological specificity, and the occurrence of inadequate subject inferences. The research, applied and exploratory in nature, adopted a qualitative and quantitative approach and was structured as a comparative subject representation experiment. The corpus comprised 16 information objects, 8 textual and 8 non-textual, from which 174 term occurrences were analyzed. The procedures involved identifying subjects in natural language, terminological and semantic validation supported by specialized glossaries, comparison with a controlled vocabulary, and analysis of terms generated by an artificial intelligence tool. The results indicated greater adherence between natural language and specialized glossaries than between natural language and the controlled vocabulary, limitations of the latter in representing emerging concepts, and the potential of artificial intelligence as a support resource, although divergences and unverified inferences were identified. It is concluded that, in the Digital Government domain, the articulation between natural language, controlled vocabulary, and artificial intelligence, combined with terminological updating and human validation, can contribute to a more consistent subject representation.
Keywords: thematic representation; knowledge organization systems; vocabulary control; electronic government; artificial intelligence.

1 INTRODUÇÃO

A representação temática constitui atividade central da Ciência da Informação (CI) e da Organização do Conhecimento (OC), uma vez que possibilita identificar, descrever e representar os assuntos presentes em diferentes objetos informacionais, contribuindo para sua organização, armazenamento, recuperação e uso. Essa atividade tem início na análise de assunto, que compreende a identificação e a seleção dos conceitos do conteúdo e sua posterior tradução em descritores, considerando o grau de especificidade, as relações semânticas e a adequação da representação às necessidades dos usuários (Campos, 2006 *apud* Dias, 2020; Hjørland, 2002; Lancaster, 2004; Gil Leiva; Fujita, 2012).

No contexto do Governo Digital, esses desafios tornam-se mais complexos em razão da articulação entre normas e estratégias, serviços públicos digitais, plataformas, dados pessoais, identidade digital, segurança da informação e Inteligência Artificial (IA). A essa complexidade soma-se a diversidade de objetos multimodais, como vídeos, infográficos, tutoriais e interfaces digitais, que precisam ser representados e recuperados (Brasil, 2021, 2024a). Essa multiplicidade de formatos, conceitos e tecnologias tensiona os instrumentos

tradicionais de representação temática, sobretudo quando os Vocabulários Controlados (VC) precisam estruturar semanticamente fenômenos em constante transformação.

Nas últimas décadas, os Sistemas de Organização do Conhecimento (SOC) passaram a coexistir com novas tecnologias e ferramentas computacionais de apoio à representação, à organização e à recuperação da informação. Mais recentemente, ferramentas baseadas em inteligência artificial e técnicas de indexação ampliaram as possibilidades de identificação e representação de assuntos em diferentes ambientes informacionais (Fujita; Sousa, 2025; Golub, 2021; Suominen, 2022). Nesse cenário, a articulação entre os recursos tecnológicos e os SOC torna-se particularmente estratégica. Os vocabulários controlados, instrumentos centrais para o controle da variabilidade terminológica inerente à linguagem natural (LN), auxiliam a minimizar problemas relacionados à polissemia e à ambiguidade, conferindo maior consistência à representação dos conteúdos (Lancaster, 2004; Lopes, 2002). Paralelamente, estratégias de indexação automática e semiautomática ampliam a capacidade de processamento de grandes volumes de objetos informacionais, sobretudo em ambientes digitais, como enfatiza Dias (2020, p. 141).

Sob essa perspectiva, a incorporação de novas tecnologias aos processos de representação temática amplia as possibilidades de automação e abre novos caminhos para a identificação e a representação dos conteúdos. Entretanto, esse uso exige uma análise crítica de suas potencialidades, limitações e implicações para a organização e a recuperação da informação, sobretudo diante dos riscos relacionados à precisão terminológica, à consistência dos resultados, à transparência dos critérios adotados e à geração de inferências não verificadas ou semanticamente inadequadas (Fujita; Sousa, 2025; Ji *et al.*, 2023).

Diante desse cenário, este estudo orienta-se pela seguinte questão: como se diferenciam as representações temáticas de objetos informacionais textuais e não textuais do domínio do Governo Digital quando produzidas a partir de termos livres da linguagem natural, de vocabulários controlados e de uma ferramenta de inteligência artificial?

Para responder a essa questão, o estudo teve como objetivo analisar comparativamente os resultados da representação temática de objetos informacionais textuais e não textuais do domínio do Governo Digital, produzidos com base em termos livres da linguagem natural, do vocabulário controlado e de uma ferramenta de inteligência artificial, considerando três critérios: a aderência semântica, a especificidade dos termos e a incidência de inferências temáticas.

A investigação resulta da adaptação de um experimento de representação temática comparada, aplicado a 16 objetos informacionais relacionados ao domínio do Governo Digital. O experimento foi desenvolvido colaborativamente entre por quatro pesquisadores, sendo um bibliotecário com experiência em representação temática e três pesquisadores com atuação profissional no domínio do Governo Digital. Com isso, o estudo adota uma perspectiva interdisciplinar que articula conhecimentos da OC, da representação da informação e das políticas e tecnologias digitais governamentais. Essa articulação possibilitou a compreensão das especificidades conceituais e terminológicas do domínio analisado, bem como a interpretação crítica das relações semânticas estabelecidas entre os termos atribuídos por LN, os descritores provenientes de VC e os termos gerados por IA. Assim, a investigação ultrapassa uma perspectiva de comparação estritamente técnica entre diferentes estratégias, instrumentos e tecnologias de representação, incorporando uma análise documentária e semântica dos resultados obtidos.

Uma das contribuições deste estudo reside na realização de uma comparação empírica entre LN, VC e IA aplicada simultaneamente à representação temática de objetos informacionais textuais e não textuais do domínio do Governo Digital. Ao analisar convergências, divergências, relações de complementaridade e lacunas entre essas abordagens, a pesquisa amplia a compreensão sobre os desafios contemporâneos da OC diante da crescente integração entre o controle terminológico, a diversidade de objetos informacionais, sobretudo os multimodais, e os sistemas baseados em inteligência artificial. Evidencia, ainda, o alcance e os limites de cada abordagem no que se refere à representação e a recuperação da informação em contextos digitais.

2 REFERENCIAL TEÓRICO

Conforme delineado na introdução, a representação temática compreende duas etapas interdependentes: a análise de assunto, que consiste na identificação e na seleção dos conceitos que representam o conteúdo de um objeto informacional, e a tradução, que converte esses conceitos em termos ou descritores que possibilitem sua organização e recuperação (Fujita, 2004; Lancaster, 2004). A primeira etapa é de natureza cognitiva e interpretativa, pois depende da compreensão do domínio ao qual o objeto informacional pertence; a segunda é de natureza terminológica e instrumental, pois depende da consistência dos SOC empregados (Dahlberg, 1978; Hjørland, 2002). A eficácia da representação, portanto,

depende de ambos os fatores. No contexto do Governo Digital, novos conceitos, tecnologias e práticas informacionais são constantemente incorporados, o que impõe desafios à atualização e à adequação dos processos e dos instrumentos de representação temática (Redígolo; Fujita, 2015).

Entre as formas de representação temática, a linguagem natural destaca-se por sua flexibilidade, pois se aproxima da forma como os usuários expressam suas necessidades informacionais e favorece a incorporação de conceitos ainda não formalizados em SOC. Entretanto, a ausência de controle terminológico pode resultar em ambiguidades, sinonímias, polissemia e variações linguísticas, comprometendo a consistência da representação e a precisão da recuperação da informação (Fujita, 2004; Lancaster, 2004; Lopes, 2002).

Em contrapartida, os vocabulários controlados promovem a padronização terminológica por meio da seleção de termos autorizados e da explicitação de relações semânticas, contribuindo para a interoperabilidade entre sistemas e para a recuperação consistente da informação (Lancaster, 2004). Neste estudo, utiliza-se o Vocabulário Controlado do Governo Eletrônico (VCGE), desenvolvido para apoiar a organização e o acesso à informação no âmbito governamental (Brasil, 2016). Apesar de sua relevância institucional, estudos de Gama e Rodrigues (2018) e Souza e Malin (2021) apontam limitações na representação de conteúdos especializados e emergentes do domínio governamental.

Mais recentemente, as ferramentas de IA introduziram novas possibilidades para a representação temática, fundamentadas na identificação de padrões e regularidades em grandes volumes de dados e no emprego de técnicas de aprendizagem de máquina para a análise e a atribuição automática de assuntos. Embora esses processos possam reduzir a dependência de procedimentos baseados exclusivamente em regras e em correspondências terminológicas previamente estabelecidas, os vocabulários controlados continuam a desempenhar um papel relevante nos sistemas de indexação assistida por IA: fornecem um repertório terminológico estável, garantem consistência e interoperabilidade entre sistemas e servem de referência para a validação dos termos atribuídos automaticamente (Fujita; Sousa, 2025; Golub, 2021).

Esses recursos ampliam a capacidade das ferramentas de inteligência artificial de contextualizar conteúdos e identificar relações semânticas, favorecendo aplicações voltadas ao apoio à indexação e à classificação documental (; Fujita, 2025; Suominen, 2022). Entretanto, como demonstram Ji *et al.* (2023), a natureza probabilística dos sistemas de IA os

torna suscetíveis à ocorrência de alucinações, que podem se manifestar na forma de imprecisões terminológicas, generalizações excessivas e associações conceituais equivocadas. Diante desse cenário, torna-se não apenas recomendável, mas necessário submeter os resultados gerados por esses sistemas a uma rigorosa validação humana e articulá-los a repertórios terminológicos consolidados.

3 PROCEDIMENTOS METODOLÓGICOS

Neste estudo, a expressão linguagem natural refere-se aos termos livres identificados pelos pesquisadores durante a análise de assunto, que constituíram a base da indexação livre. A pesquisa caracteriza-se como aplicada e exploratória, com abordagem qualitativa e quantitativa, estruturando-se como um experimento de representação temática comparada. Foram analisadas três abordagens de representação temática: a indexação livre, baseada na linguagem natural; a indexação controlada, baseada em descritores provenientes do VCGE; e a indexação assistida por IA, baseada em termos gerados pelo Gemini. Como instrumentos complementares, foram utilizados glossários especializados do domínio do Governo Digital, com o propósito de apoiar a validação terminológica e semântica dos termos identificados. A dimensão qualitativa corresponde à interpretação das relações semânticas estabelecidas entre os termos, enquanto a dimensão quantitativa consiste na contagem das ocorrências em cada categoria de aderência terminológica.

Para viabilizar essa análise, o corpus foi composto por 16 objetos informacionais relacionados ao domínio do Governo Digital brasileiro, 8 textuais e 8 não textuais, selecionados intencionalmente com base em critérios de pertinência temática, origem em fontes oficiais, diversidade tipológica, presença de conceitos emergentes, além de relevância para temas como serviços públicos digitais, identidade digital, proteção de dados pessoais, segurança da informação e IA. Os objetos textuais incluíram documentos estratégicos e normativos (leis, decretos e portarias); os objetos não textuais abrangeram vídeos institucionais, tutoriais, além de materiais de divulgação e imagens informativas.

O VCGE (Brasil, 2016) foi adotado como o principal instrumento de representação temática e como referência para análise comparativa. Contudo, estudos como os de Gama e Rodrigues (2018) e Souza e Malin (2021) apontam limitações quanto à capacidade desse instrumento de representar temas especializados e emergentes relacionados à transformação digital do Estado. Por essa razão, recorreu-se ao uso complementar de glossários

especializados como fontes de apoio terminológico e semântico, a saber: o *Glossário de Segurança da Informação* (Brasil, 2021), o *Glossário de Termos de Dados* (Brasil, 2024), o *Glossário de Termos Relacionados à Inteligência Artificial* (Brasil, 2025a) e o *Glossário de Termos Relacionados à IA Generativa* (Brasil, 2025b). Tais instrumentos apoiaram a delimitação e a validação dos conceitos identificados no corpus.

A ferramenta de IA utilizada no experimento foi o Gemini 3.5 Flash, acessado em sua versão gratuita. Os prompts foram estruturados para orientar a identificação do tema central, dos assuntos secundários, dos conceitos e das palavras-chave, bem como de termos correspondentes aos descritores do VCGE. No caso de objetos não textuais, os prompts incluíram orientações para a descrição dos elementos visuais e para a distinção entre informações diretamente observáveis e inferências contextuais.

As cinco etapas descritas a seguir correspondem aos procedimentos analíticos do experimento e não à quantidade de estratégias de representação temática analisadas. A **primeira etapa** consistiu na análise de assunto dos 16 objetos e na identificação de termos em linguagem natural pelos pesquisadores, responsáveis pela indexação. Como resultado desse procedimento, foram identificadas 174 ocorrências de termos livres, que constituíram a base para as comparações realizadas nas etapas subsequentes. Em síntese, os 16 objetos correspondem ao corpus documental da pesquisa, enquanto as 174 ocorrências constituem as unidades de análise empregadas nas comparações terminológicas e semânticas. Na **segunda etapa,** os glossários especializados foram utilizados como instrumentos de apoio à validação dos termos identificados na etapa anterior. Para isso, os termos livres, atribuídos em LN foram confrontados com os termos e as definições disponíveis nos glossários, a fim de verificar a existência e o grau de correspondência terminológica entre eles. Na **terceira etapa,** os termos com definições validadas pelos glossários, foram comparados com os descritores disponíveis no VCGE. Sempre que necessário e possível, realizou-se a tradução desses termos para os descritores disponíveis no VCGE. Na **quarta etapa,** os resultados da indexação realizada com base no VCGE foram comparados com os termos sugeridos pelo Gemini. Para essa finalidade, a ferramenta foi orientada, por meio de um prompt previamente elaborado, a identificar os assuntos dos objetos e a utilizar o VCGE como referência para a atribuição dos termos de representação temática. A comparação permitiu analisar os níveis de correspondência e as divergências entre a indexação fundamentada no vocabulário controlado e os resultados produzidos pela ferramenta. Por fim, a **quinta etapa** consistiu na

comparação entre os resultados da representação temática de objetos textuais e não textuais, realizados pelo Gemini, com o objetivo de identificar diferenças relacionadas às características e às especificidades de cada tipo de objeto informacional.

A identificação de 174 ocorrências de termos livres constituiu a base empírica das comparações realizadas nas etapas subsequentes. Cada uma delas foi considerada uma unidade de análise nas comparações terminológicas e semânticas entre os termos livres, os descritores do VCGE e os termos sugeridos pelo Gemini, tendo os glossários como instrumento de apoio à validação. Quanto à aderência semântica, foram adotadas as categorias analíticas apresentadas no Quadro 1.

Quadro 1 - Categorias utilizadas na análise de aderência semântica

Tipo	**Etapa de Aplicação**		**Categoria**	**Definição analítica**
Transversal	**2ª, 3ª e 4ª**	**SC**	Sem Correspondência	Não foi identificado termo ou descritor capaz de representar, de forma satisfatória ou aproximada, o conceito analisado.
		CE	Correspondência Exata	Mesmo termo (forma lexical) e mesmo conceito (mesma definição)
Específica	**2ª**	**CA**	Correspondência Aproximada	Os termos representam conceitos semanticamente próximos, mas não equivalentes, apresentando correspondência parcial.
		CC	Correspondência Conceitual	Termos diferentes, mas mesmo conceito
Específica	**3ª**	**TR**	Termo Relacionado	Existe uma relação semântica pertinente entre os termos, sem que haja equivalência conceitual ou uma relação hierárquica direta.
		TG	Termo Genérico	O termo encontrado representa um conceito mais amplo do que o termo ou descritor com o qual foi comparado
Específica	**4ª**	**DV**	Divergência	Termo existente, identificável no universo terminológico considerado (VCGE), sugerido pela IA, mas inadequado (fora do escopo) para representar o assunto específico do objeto.
		AL	Alucinação	Termo ou conceito gerado pela IA sem sustentação verificável no instrumento terminológico utilizado como referência (VCGE).

Fonte: elaboração própria (2026).

Os resultados foram organizados em tabelas de ocorrência e interpretados de acordo com cada proposta de comparação. A dimensão quantitativa permitiu identificar a distribuição das categorias utilizadas em cada etapa, enquanto a dimensão qualitativa examinou o significado dessas ocorrências, especialmente em relação às categorias CE, CC, CA, TG, TR, SC, DV e AL, descritas no Quadro 1. Nesse sentido, os percentuais e as frequências não foram interpretados isoladamente, mas articulados às características dos termos, dos instrumentos e dos objetos informacionais analisados.

4 RESULTADOS E DISCUSSÃO

Esta seção apresenta e discute os resultados obtidos a partir das análises comparativas entre as três abordagens de representação temática analisadas, indexação livre, indexação controlada e indexação assistida por IA, destacando tanto as convergências e divergências entre elas quanto a suas potencialidades e limitações na representação de objetos informacionais textuais e não textuais do domínio do Governo Digital.

4.1 LINGUAGEM NATURAL E GLOSSÁRIOS ESPECIALIZADOS

A primeira comparação corresponde à segunda etapa metodológica, a validação terminológica e semântica dos termos identificados em linguagem natural, realizada com o apoio de glossários especializados. Os resultados apresentados na Tabela 1 indicam que 54,6% das 174 ocorrências analisadas (ver seção 3) apresentaram algum nível de correspondência semântica, CE e CC (18,4 %) e CA (36,2%). A categoria SC alcançou 45,4%, evidenciando lacunas na cobertura terminológica dos glossários em relação ao corpus analisado.

Tabela 1 – Aderência semântica entre linguagem natural e Glossários

Tipo de resultado	**Categorias**	**Ocorrências**	**%**	**% acumulado**
Correspondência satisfatória	CE + CC	32	18,4%	54,6%
Cobertura parcial	CA	63	36,2%	
Ausência de cobertura	SC	79		45,4%
Total	—	**174**		**100%**

Fonte: elaboração própria (2026)

Após essa etapa, a análise avançou para a comparação entre a LN e o VCGE.

2 LINGUAGEM NATURAL E VOCABULÁRIO CONTROLADO DO GOVERNO ELETRÔNICO

A comparação entre a linguagem natural e o Vocabulário Controlado do Governo Eletrônico revelou que apenas duas ocorrências (1,1%) foram classificadas correspondências exatas (CE) e nenhuma como correspondência conceitual com termos diferentes (CC), o que indica um cenário de aderência semântica predominantemente indireta e limitada. Das 174 ocorrências analisadas, apenas duas (1,1%) foram classificadas como CE, evidenciando a baixa incidência de padronização terminológica entre os termos da linguagem natural e os descritores do vocabulário controlado. A categoria predominante foi SC, com 63 ocorrências (36,2%), indicando que uma parcela significativa dos termos atribuídos aos objetos analisados não encontrou representação no VCGE. Em seguida, destacaram-se os TR, com 57 ocorrências (32,8%) e os TG, com 52 ocorrências (29,9%), conforme apresentado na Tabela 2. Esse resultado sugere que o VCGE, embora útil para a representação de categorias gerais, apresenta limitações para representar conceitos específicos e contemporâneos do domínio do Governo Digital.

Tabela 2 – Aderência semântica entre linguagem natural e vocabulário controlado

	Categoria	**Ocorrências**	**%**
SC	Sem Correspondência	63	36,2%
TR	Termo Relacionado	57	32,8%
TG	Termo Genérico	52	29,9%
CE	Correspondência Exata	2	1,1%
CC	Correspondência Conceitual	0	0,0%
	Total	**174**	**100%**

Fonte: elaboração própria (2026)

O dado mais relevante, portanto, não se restringe à reduzida ocorrência de correspondências exatas, mas refere-se à própria natureza da lacuna identificada, uma vez que não foi registrada pelo menos correspondência conceitual entre os termos da linguagem natural e os do vocabulário controlado. Embora o vocabulário se mostre útil para a representação de categorias gerais da administração pública, como administração, serviços públicos, planejamento e gestão, comunicações e cadastro, sua estrutura apresenta menor aderência a conceitos mais específicos e contemporâneos relacionados ao Governo Digital, especialmente aqueles vinculados a plataformas digitais, dados, autenticação, segurança cibernética e inteligência artificial.

4.3 INDEXAÇÃO BASEADA NO VCGE E TERMOS SUGERIDOS PELO GEMINI

A quarta comparação analisou a aderência semântica entre os descritores do VCGE atribuídos pelo grupo de indexadores e os termos gerados pelo Gemini, no processo de indexação assistida por IA, baseada no próprio VCGE. Os resultados evidenciam 89 correspondências exatas (CE), que representam 51,1% dos 174 registros analisados, e 76 divergências (DV), equivalentes a 43,7%. Foram identificadas, ainda, cinco ocorrências sem correspondência (SC) (2,9%) e quatro alucinações (AL) (2,3%).

Os resultados evidenciam um desempenho heterogêneo da IA na representação temática. Embora 51,1% das ocorrências tenham sido classificadas como correspondência exata, esse resultado coexistiu com uma proporção expressiva de divergências (43,7%), indicando que a ferramenta não apresentou consistência suficiente para utilização autônoma no contexto analisado. Assim, os dados evidenciam um comportamento heterogêneo, caracterizado pela capacidade da ferramenta de produzir termos coincidentes com os descritores atribuídos pelos indexadores. No entanto, essa capacidade coexiste com uma proporção significativa de resultados que não correspondem adequadamente à representação temática considerada pertinente.

As categorias SC (2,9%) e AL (2,3%), embora menos frequentes, também revelam limitações na geração automática de termos. Em conjunto, os resultados indicam que a ferramenta apresenta potencial como recurso de apoio à representação temática, mas não demonstra consistência suficiente para sua utilização autônoma. Assim, a aplicação da IA nesse processo requer validação humana, articulação com critérios documentários e utilização de instrumentos de controle terminológico.

Conforme definido no Quadro 1, a categoria DV aplica-se a um termo sugerido pela inteligência artificial que, embora relacionado ao contexto do objeto, não representa adequadamente o assunto ou o conceito pertinente à sua representação temática. A categoria AL, por sua vez, refere-se a um termo ou conceito gerado pela inteligência artificial sem sustentação verificável no instrumento terminológico utilizado como referência. Embora tenham sido identificadas apenas quatro ocorrências dessa categoria, sua relevância metodológica não deve ser minimizada, pois evidenciam a possibilidade de a IA produzir representações aparentemente pertinentes que não atendem aos critérios terminológicos estabelecidos. Esses resultados reforçam, portanto, a necessidade de validação humana das

sugestões geradas automaticamente e da articulação da IA com instrumentos de controle terminológico.

Tabela 3 – Aderência semântica entre VCGE e Gemini

	Categoria	**Ocorrências**	**%**
CE	Correspondência Exata	89	51,1%
DV	Divergência	76	43,7%
SC	Sem Correspondência	5	2,9%
AL	Alucinação	4	2,3%
	Total	**174**	**100%**

Fonte: elaboração própria (2026).

Os resultados apresentados permitem avançar para a análise das convergências e divergências na representação temática de objetos textuais e não textuais.

4.4 OBJETOS TEXTUAIS E NÃO TEXTUAIS

A quinta comparação analisou os resultados da representação temática de objetos textuais e não textuais, com o apoio do Vocabulário Controlado do Governo Eletrônico e do Gemini. Esta seção teve como objetivo identificar diferenças no desempenho dessas abordagens de representação em função das características e das especificidades de cada tipo de objeto informacional.

Os objetos textuais, predominantemente de natureza normativa, estratégica e institucional, apresentaram maior estabilidade semântica, uma vez que os conceitos tendem a ser explicitados diretamente no conteúdo, favorecendo uma maior aproximação entre os termos identificados na linguagem natural, os descritores do vocabulário controlado e as inferências produzidas pela ferramenta de IA. A análise qualitativa dos resultados sugere maior estabilidade semântica nos objetos textuais e maior dispersão nos objetos não textuais. Vídeos, tutoriais, materiais de divulgação e imagens informativas ampliaram a incidência de inferências contextuais e de associações conceituais equivocadas. O Gemini demonstrou capacidade de descrever elementos visuais e contextualizar os objetos, mas nem sempre distinguiu adequadamente o que era elemento visual observável, significado institucional e descritor temático. Essa dificuldade reforça a necessidade de protocolos específicos para representação temática de objetos não textuais.

Os resultados sugerem que a representação temática de objetos não textuais exige um nível adicional de mediação profissional. Em vez de aceitar diretamente os descritores gerados

por IA, o profissional da informação deve validar se o termo representa o assunto do objeto ou apenas um elemento visual periférico. Essa distinção é decisiva para evitar ruído na recuperação da informação. O Quadro 2 sintetiza as principais diferenças observadas entre objetos textuais e não textuais.

Quadro 2 - Síntese comparativa entre objetos textuais e não textuais

Categoria	Objetos textuais	Objetos não textuais
Características gerais	Maior estabilidade semântica.	Maior dispersão semântica.
Relações semânticas predominantes	Maior ocorrência de **TG** (termos genéricos) e **TR** (termos relacionados).	Maior incidência de **SC** (sem correspondência) e **AL** (alucinações).
Comportamento da ferramenta de IA	Menor incidência de **inferências não verificadas**.	**Inferências contextuais** mais amplas.
Relação com o VCGE	Maior proximidade com categorias administrativas e normativas.	Dificuldade de correspondência direta com elementos visuais.
Risco principal	Perda de especificidade por generalização.	Confusão entre descrição visual e representação temática.

Fonte: elaboração própria (2026).

À luz das reflexões de Dias (2020), as discrepâncias observadas na representação temática de objetos textuais e não textuais não devem ser atribuídas exclusivamente às potencialidades e limitações das abordagens empregadas. Tais variações também podem decorrer das características intrínsecas aos próprios objetos informacionais e dos distintos modos pelos quais seus conteúdos, conceitos e significados são identificados, interpretados e representados.

4.5 SÍNTESE CRÍTICA DAS ESTRATÉGIAS DE REPRESENTAÇÃO

Os resultados permitem interpretar as três abordagens como complementares, não substitutivas. A indexação livre mostrou maior proximidade com o vocabulário dos usuários e maior capacidade de nomear conceitos emergentes. A principal limitação dessa abordagem reside na ausência de controle terminológico, que favorece ambiguidades e reduz a consistência da representação.

A indexação controlada, baseada no Vocabulário Controlado do Governo Eletrônico demonstrou valor para padronização, interoperabilidade e organização institucional da informação. Contudo, apresentou baixa especificidade para temas tecnológicos emergentes.

A limitação do VCGE não invalida o instrumento; indica necessidade de atualização, articulação com glossários especializados e possível evolução para estruturas terminológicas mais dinâmicas.

A indexação assistida por IA, apresentou potencial para a indexação, principalmente na identificação de relações contextuais e no tratamento inicial de objetos não textuais. Porém, os resultados indicam que sua utilização isolada é inadequada. A ferramenta deve ser utilizada em conformidade com critérios documentários, processos de validação humana e mecanismos de controle terminológico, sob pena de produzir termos plausíveis, mas sem validade terminológica para recuperação consistente da informação.

Em conjunto, os resultados indicam que a representação temática de objetos informacionais do Governo Digital exige a articulação das três abordagens — indexação livre, controlada e indexação assistida por IA —, uma vez que cada uma cobre lacunas das demais. As diferenças observadas não indicam superioridade de uma sobre as outras, mas distintos alcances e limitações que as tornam complementares.

5 CONSIDERAÇÕES FINAIS

Este experimento analisou comparativamente abordagens de representação temática baseadas em linguagem natural, vocabulário controlado e inteligência artificial aplicadas a objetos informacionais textuais e não textuais do domínio de Governo Digital. A comparação das 174 ocorrências permitiu identificar diferenças relativas à aderência semântica, à especificidade terminológica e à ocorrência de inferências inadequadas, respondendo ao objetivo proposto. A questão-problema foi respondida ao demonstrar que as principais diferenças entre as abordagens decorrem de seus distintos mecanismos de identificação e controle dos assuntos: a linguagem natural apresenta maior flexibilidade terminológica; o vocabulário controlado favorece a padronização, mas apresenta lacunas diante de conceitos emergentes; e a inteligência artificial amplia a contextualização, mas produz divergências e inferências que requerem validação humana.

A análise da linguagem natural evidenciou sua capacidade de representar conceitos emergentes e de acompanhar as transformações terminológicas próprias do domínio do Governo Digital. A validação dos termos por meio dos glossários revelou, contudo, limitações desses instrumentos para acompanhar a dinâmica e constante evolução terminológica da área. Os resultados quantitativos reforçam essa interpretação: a comparação com os

glossários especializados (Tabela 1) apresentou 54,6% de correspondências satisfatórias, evidenciando aderência relevante, mas também 45,4% de ausência de correspondência. O VCGE, por sua vez, apresentou potencial de padronização e interoperabilidade no contexto da Administração Pública, mas revelou baixa aderência a conceitos contemporâneos relacionados ao Governo Digital, à inteligência artificial, à segurança cibernética e à identidade digital. Conforme demonstrado na Tabela 2, a elevada ocorrência da categoria SC (Sem Correspondência), a baixa incidência de CE (Correspondência Exata) e a ausência de CC (Correspondência Conceitual) expressam essa limitação e sustentam qualitativamente a necessidade de atualização terminológica do Vocabulário Controlado do Governo Eletrônico.

Os resultados da indexação assistida por IA revelaram um cenário heterogêneo, caracterizado pela coexistência de elevada ocorrência de correspondências exatas e de um percentual expressivo de divergências. Embora 51,1% das ocorrências tenham apresentado CE com os descritores do VCGE atribuídos pelos indexadores, 43,7% foram classificadas como DV, evidenciando limitações relevantes quanto à precisão e à preservação da especificidade conceitual. Os casos de SC (2,9%) e de AL (2,3%), ainda que menos frequentes, também indicam a ocorrência de resultados que não atendem plenamente aos critérios terminológicos estabelecidos.

Para a Organização do Conhecimento, o estudo contribui ao evidenciar, com base em resultados quantitativos articulados à análise qualitativa, que a linguagem natural, o vocabulário controlado e a inteligência artificial desempenham funções complementares na representação temática de objetos informacionais em ambientes digitais. Os percentuais de aderência e divergência demonstram que nenhuma das abordagens, isoladamente, respondeu de forma suficiente às necessidades de representação do corpus analisado. Assim, a crescente presença de objetos não textuais e a utilização de ferramentas baseadas em inteligência artificial não reduzem a importância do controle terminológico e da mediação humana; ao contrário, reforçam a necessidade de abordagens capazes de articular padronização conceitual, contextualização semântica e atualização contínua dos SOC.

Conclui-se que a inteligência artificial pode apoiar processos de representação temática, sobretudo como recurso de pré-indexação, contextualização e análise inicial de objetos não textuais. Contudo, sua adoção exige supervisão humana, critérios metodológicos explícitos e controle terminológico. A análise realizada demonstra que a complementaridade entre linguagem natural, Sistemas de Organização do Conhecimento e inteligência artificial

constitui uma alternativa mais consistente do que a substituição de uma estratégia por outra. Nesse sentido, os resultados reforçam a centralidade da mediação especializada nos processos de Organização do Conhecimento, uma vez que a avaliação das relações conceituais, das equivalências terminológicas e da adequação semântica da representação ultrapassa capacidades estritamente automatizadas.

Como agenda futura, recomenda-se ampliar o corpus, testar diferentes ferramentas de inteligência artificial, comparar versões gratuitas e pagas, aplicar métricas de precisão e revocação em sistemas reais de recuperação da informação e investigar estratégias para atualização contínua dos vocabulários controlados, incluindo os glossários especializados neste domínio e o próprio Vocabulário Controlado do Governo Eletrônico, com base nas garantias literária, cultural e dos usuários, bem como em evidências de uso provenientes dos serviços públicos digitais.

REFERÊNCIAS

BRASIL. Ministério do Planejamento, Orçamento e Gestão. Secretaria de Logística e Tecnologia da Informação. **VCGE:** Vocabulário Controlado do Governo Eletrônico. Versão 2016, Número 2.1.0. Tipo principal. Documento detalhado. Brasília, DF: MP/SLTI, 2016. 93 p. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/arquivos/vcge_2_1_0_det.pdf. Acesso em: 11 maio 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. **Glossário de termos de dados**. Brasília, DF: MGI, 2024a. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/governancadedados/glossario-de-termos-de-dados . Acesso em: 16 maio 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **Glossário de termos relacionados à inteligência artificial**. Brasília, DF: MGI, 2025a. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/inteligencia-artificial-1/publicacoes/glossario-de-termos-relacionados-a-ia . Acesso em: 16 maio 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. **Inteligência artificial generativa no serviço público**: glossário de termos relacionados à inteligência artificial. Brasília, DF: Governo Digital, 2025b. Disponível em: https://www.gov.br/governodigital/pt-br/infraestrutura-nacional-de-dados/inteligencia-artificial-1/glossario-ia-generativa-no-servico-publico.pdf. Acesso em: 13 maio 2026.

BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. Portaria SGD/MGI nº 4.248, de 26 de junho de 2024b. Estabelece recomendações para o alcance dos objetivos da Estratégia Nacional de Governo Digital para o período de 2024 a 2027. **Diário Oficial da União:** seção 1, Brasília, DF, n. 123, p. 111, 28 jun. 2024.

Disponível em: https://www.in.gov.br/en/web/dou/-/portaria-sgd/mgi-n-4.248-de-26-de-junho-de-2024-568659997 . Acesso em: 24 maio 2026.

BRASIL. Gabinete de Segurança Institucional da Presidência da República. Portaria GSI/PR nº 93, de 18 de outubro de 2021. Aprova o glossário de segurança da informação. **Diário Oficial da União:** seção 1, Brasília, DF, p. 36, 19 out. 2021. Disponível em: https://www.in.gov.br/en/web/dou/-/portaria-gsi/pr-n-93-de-18-de-outubro-de-2021-353056370. Acesso em: 11 maio 2026.

DAHLBERG, Ingetraut. A referent-oriented, analytical concept theory for INTERCONCEPT. **International Classification**, Frankfurt, v. 5, n. 3, p. 142-151, 1978. Disponível em: https://storage.imrpress.com/imr/journal/KO/article/506826/1752851142083.pdf. Acesso em: 11 maio 2026.

DIAS, Célia da Consolação. Representação temática de imagens: reflexões acerca dos subsídios da indexação manual e do reconhecimento de imagens. **Perspectivas em Ciência da Informação,** Belo Horizonte, v. 25, número especial, p. 125-149, fev. 2020. Disponível em: https://hdl.handle.net/1843/37009. Acesso em: 15 maio 2026.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, 2004. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089. Acesso em: 29 maio 2026.

FUJITA, Mariângela Spotti Lopes; SOUSA, Nuno Miguel Teixeira. Vocabulário controlado e inteligência artificial na indexação: uma revisão bibliográfica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 30, 2025. Disponível em: https://www.scielo.br/j/pci/a/br3MytQmVck88FTJPgTJvvC/. Acesso em: 19 maio 2026.

GAMA, Janyluce Rezende; RODRIGUES, Georgete Medleg. O vocabulário controlado do Governo Eletrônico: contribuições e limites na implementação da Lei de Acesso à Informação no Brasil. **Em Questão**, Porto Alegre, v. 24, n. 1, p. 94-119, jan./abr. 2018. Disponível em: https://seer.ufrgs.br/index.php/emquestao/article/view/70989 . Acesso em: 29 maio 2026.

GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes (ed.). **Política de indexação**. Marí lia: Oficina Universitária, 2012. 260p.

GOLUB, Koraljka. Automated Subject Indexing: An Overview. **Cataloging & Classification Quarterly,** [*s. l.*], v. 59, n. 8, p. 702–719, 2021. Disponível em: https://www.tandfonline.com/doi/pdf/10.1080/01639374.2021.2012311. Acesso em: 26 maio 2026.

HJØRLAND, Birger. Epistemology and the socio-cognitive perspective in information science. **Journal of the American Society for Information Science and Technology,** New York, v. 53,

n. 4, p. 257- 270, 2002. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/asi.10042. Acesso em: 29 maio 2026.

JI, Ziwei *et al.* Survey of hallucination in natural language generation. **ACM Computing Surveys,** New York, v. 55, n. 12, art. 248, p. 1-38, 2023. Disponível em: https://dl.acm.org/doi/epdf/10.1145/3571730. Acesso em: 29 maio 2026.

LANCASTER, Frederick Wilfrid. **Indexação e resumos:** teoria e prática. Tradução Antonio Agenor Briquet de Lemos. 2. ed. Brasília**,** DF: Briquet de Lemos, 2004. 452 p.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ciência da Informação**, Brasília, DF, v. 31, n. 1, p. 41-52, jan./abr. 2002. Disponível em: https://www.scielo.br/j/ci/a/WhYch5gHnWYPCPY3NPwbkcR/. Acesso em: 17 maio 2026.

REDÍGOLO, Franciele Marques; FUJITA, Mariângela Spotti Lopes. A leitura profissional do catalogador de assunto na perspectiva sociocognitiva. **Informação & Informação**, Londrina, v. 20, n. 3, p. 356-376, set./dez. 2015. Disponível em: https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/21683. Acesso em: 11 maio 2026.

SOUZA, Miriely da Silva; MALIN, Ana Maria Barcellos. O Vocabulário Controlado do Governo Eletrônico como instrumento de prestação de contas da Lei de Acesso à Informação. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 21., 2021, Rio de Janeiro. **Anais** [...]. São Paulo: ANCIB, 2021. Disponível em: https://enancib.ancib.org/enancib/pt_BR/article/view/679/691 . Acesso em: 15 maio 2026.

SUOMINEN, Osma. Developing and implementing automated subject indexing: the Annif open source toolkit. **JLIS.it**, [*s. l.*], v. 13, n. 1, p. 265-282, 2022. Disponível em: https://www.jlis.it/index.php/jlis/article/view/437. Acesso em: 29 maio 2026.