Overcoming Stakeholder Knowledge Gaps: A HITL-Oriented Ontology Mapping Framework for the Construction Industry

Jakob Martin[1], Timur Weilbach-Eyüboglu[1] and Vedant Girish Dalvi[1]
[1]University of Applied Sciences, Munich, Germany
E-mail(s): jakob.martin0@hm.edu, timur.weilbach-eyueboglu@hm.edu, vedant_girish.dalvi@hm.edu

Abstract: Semantic interoperability in construction and Facility Management (FM) processes is crucial for efficient information exchange, yet automated semantic matching systems often struggle with context-specific terminology and dynamic project requirements. This paper presents an extended Human-in-the-Loop (HITL) component integrated into an existing semantic matching platform for construction and FM workflows. Unlike traditional approaches with simple feedback loops, our system evolves the HITL paradigm into a role-based, domain-specific expert system. The HITL extension introduces multiple innovations: role-weighted feedback, where inputs from users (e.g., architect, engineer, facility manager) are weighted by expertise and project role; adaptive interface modes tailored to user expertise and project phase; an active learning module that detects uncertainty in match suggestions and queries experts; a conflict escalation hierarchy to resolve contradictory feedback through tiered expert intervention; and phase-specific weighting logic adjusting the matching algorithm to the current project phase. We detail the design and implementation of these features and report an evaluation of real-world construction data. Results show that the enhanced HITL system improves matching accuracy (precision 91.8%, recall 92.0%) and reduces manual corrections by 30%, demonstrating the effectiveness of integrating human expertise into semantic matching for the Architecture, Engineering, Construction and Facility Management (AEC/FM) domain.

Keywords: Human-in-the-Loop, Semantic Data Mapping, Role-Aware Interaction, Active Learning, Building Information Modeling (BIM)



DOI: 10.18154/RWTH-CONV-254909. Published in the conference proceedings of the 36. Forum Bauinformatik 2025, Aachen, Germany,
© 2025 The copyright for this article lies with the authors. This publication, except for quotations and otherwise indicated parts, is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

1 Introduction

Semantic data integration is a longstanding challenge in the Architecture, Engineering, Construction and Facility Management (AEC/FM) domain [1]. Projects involve heterogeneous data, leading to mismatches and silos. [2]. Semantic matching systems address this by automatically identifying equivalent or related concepts across different representations, for instance, matching a facility management (FM) task description to a Building Information Modelling (BIM) component. While such

systems leverage ontologies and machine learning (ML) to improve interoperability, they often fail to capture project-specific nuances and evolving domain semantics.

Human-in-the-Loop (HITL) strategies have emerged as a promising means to address these shortcomings by integrating human expertise into automated processes [3]. However, most existing HITL implementations in the AEC/FM sector are limited to simple feedback confirmation loops, offering little contextual differentiation [4]. They do not account for the varying expertise of stakeholders, project lifecycle stages, or feedback conflicts.

This paper presents an extended HITL framework that addresses these limitations by incorporating role-specific feedback weighting, adaptive user interaction, and active learning strategies into a semantic matching system for construction and FM workflows. The framework aims to align automated suggestions more closely with project realities while reducing manual correction effort.

The remainder of this paper is structured as follows: Section 2 reviews related work in semantic matching and HITL systems. Section 3 outlines the design of the extended HITL framework. Section 4 describes the implementation within a semantic BIM matching tool. Section 5 presents evaluation results and discusses key findings. Section 6 concludes the paper and suggests future work.

2 Related Work

Prior research highlights how semantic enrichment improves information exchange and analysis across heterogeneous construction data sources [2]. Despite these advances, systems remain mostly automated and often miss domain-specific nuances. Generic algorithms struggle with varied terminology and subtle distinctions, such as between ”commissioning” and ”handover.” Some frameworks support manual validation, but usually as one-time interventions rather than continuous learning loops [2], [5].

By contrast, the broader machine learning field has developed more sophisticated HITL strategies. Active learning techniques help the system identify cases with high uncertainty and request input from human users for clarification. In contrast, interactive machine learning focuses on refining the model step by step through ongoing user feedback [3]. These techniques are especially effective in domain-specific applications like knowledge graph construction [6]. Some AEC/FM tools support compliance checks or ontology editing but still rely on simple approvals and ignore structured feedback. Expert systems in construction typically encode static rules; however, modern HITL approaches offer a more dynamic alternative, incorporating expert feedback into ongoing model updates. However, role-weighted feedback remains underexplored. While access control is common, few systems differentiate the influence of input based on expertise.

Another key dimension is user interface design. Human-centered computing has shown that adaptive User-Interfaces (UIs) help balance usability for both experts and non-experts [7]. In BIM and FM systems, overloaded interfaces are a known challenge [7]. Some tools offer role-specific views, but they are rarely dynamic. Active learning and uncertainty estimation are well-established in text classification, but still emerging in construction informatics. By learning from individual feedback and modeling how consensus is formed, such systems can capture distributed domain knowledge more effectively.

3 Methodology: Extended HITL System Design

This section outlines the methodological foundation of the extended HITL framework.

The framework overcomes limitations of conventional feedback in semantic matching. Traditional approaches typically treat all user feedback as equal and static, without considering contextual factors such as user expertise, project phase, or the confidence level of a match. In contrast, our system introduces multiple mechanisms to ensure that human input is contextually weighted, intelligently solicited, and effectively applied.

Because roles vary across procurement and delivery strategies, the framework avoids a fixed hierarchy of user influence. It uses a configurable role-weighting matrix with default profiles that can be adapted to project-specific structures.

A central component of the system is the role-based feedback weighting logic. User input is evaluated in light of the contributor's domain expertise and responsibilities. For example, during the design phase, feedback from architects or domain specialists is prioritized, while in the operational phase, facility managers receive greater influence. This weighting mechanism reflects real-world trust hierarchies and aligns the system with project workflows. Probabilistic models such as Bayesian updating and learning-to-rank methods adjust future match confidence based on the weighted feedback.

Building on this, the system adapts its interaction logic. Interface modes are selected according to user profile, technical background, and behavioral patterns such as feedback frequency or correction history. Following progressive disclosure principles [8], expert users are offered detailed editing capabilities, while non-experts receive guided, step-by-step interaction modes. If a less experienced user frequently encounters low-confidence suggestions, the system automatically shifts to Guided Mode. Conversely, repetitive confirmations may activate a Passive Mode to reduce cognitive load.

To focus expert input where it matters most, the system includes an active learning module. Matches associated with low confidence or semantic ambiguity are flagged and routed to relevant users based on role and expected information gain. This ensures that human attention is directed to critical decisions and allows the model to iteratively improve in high-impact areas.

In collaborative scenarios where user feedback conflicts, a conflict resolution mechanism is activated. The system categorizes disagreements based on severity and triggers escalation workflows if necessary. Initial strategies include prompting for clarification or requesting justification. If conflicts persist, feedback is escalated to a senior expert or resolved through a structured consensus process. All feedback and decisions are logged to support transparency and continuous learning.

Finally, the system accounts for the temporal dynamics of construction and FM projects. Semantic interpretation and feedback relevance change throughout the project lifecycle. For instance, broader semantic matches may be acceptable in early design, while stricter definitions are necessary during commissioning. Accordingly, the HITL framework adapts its weighting logic, interface behavior, and match acceptance thresholds in alignment with the current project phase.

By combining these features, the extended HITL framework strengthens a Bidirectional Encoder Representations from Transformers (BERT)-based semantic matching core with expert-guided refinement.

Rather than treating feedback as uniform, the system dynamically adapts to user role, experience, and project context. This approach ensures precise calibration of similarity scores, reduces manual correction effort, and increases trust in automated suggestions.

The overall system aims to improve interoperability by leveraging structured human input in uncertain, conflict, or phase-dependent ambiguity situations. Figure 1 illustrates the feedback and decision flow of the HITL framework, showing when human intervention is triggered and how validated input feeds back into the matcher.

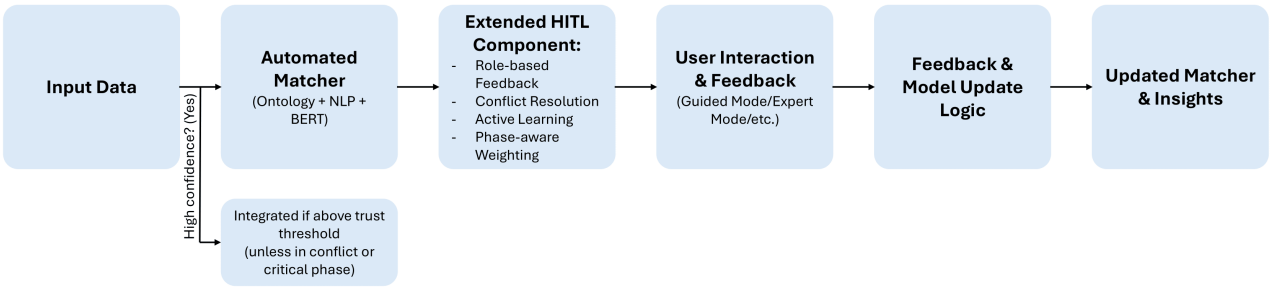


Figure 1: Processing flow of the extended HITL system. Matches with high confidence bypass human review unless flagged as critical. Ambiguous cases are routed through the HITL logic for role-aware feedback and adaptive interaction.

3.1 System Overview and Feedback Architecture

The extended HITL system builds on an existing semantic matcher that processes inputs such as process descriptions and BIM element metadata. Using ontological reasoning and natural language processing techniques, the matcher generates candidate alignments with associated confidence scores, which are then passed to the HITL module for further evaluation. These matches and associated confidence scores are passed to the HITL module for further evaluation.

The HITL module consists of an adaptive interface, a role-aware feedback weighting mechanism, and a conflict resolution system. High-confidence matches are accepted automatically, while ambiguous cases trigger human review. Feedback is evaluated based on user roles stored in a profile registry, which informs interface behavior and input weighting.

3.2 Learning Strategy and Lifecycle Adaptation

The HITL system incorporates targeted learning mechanisms and structured conflict resolution to improve semantic alignment in areas where automation proves insufficient.

The system identifies uncertain or ambiguous matches using calibrated confidence scores. Low-confidence matches are flagged and routed to qualified users. Feedback updates the model incrementally by refining similarity scores or ontology links. Prompts are embedded into task-specific interfaces at key decision points. The system is tuned to avoid overload by prioritizing high-impact interactions and suppressing redundant or low-confidence suggestions.

When users submit contradictory feedback or challenge a high-confidence suggestion, the system categorizes the disagreement by severity. Minor conflicts are resolved through clarification, while major conflicts involving users with high authority trigger an escalation procedure. In escalated cases,

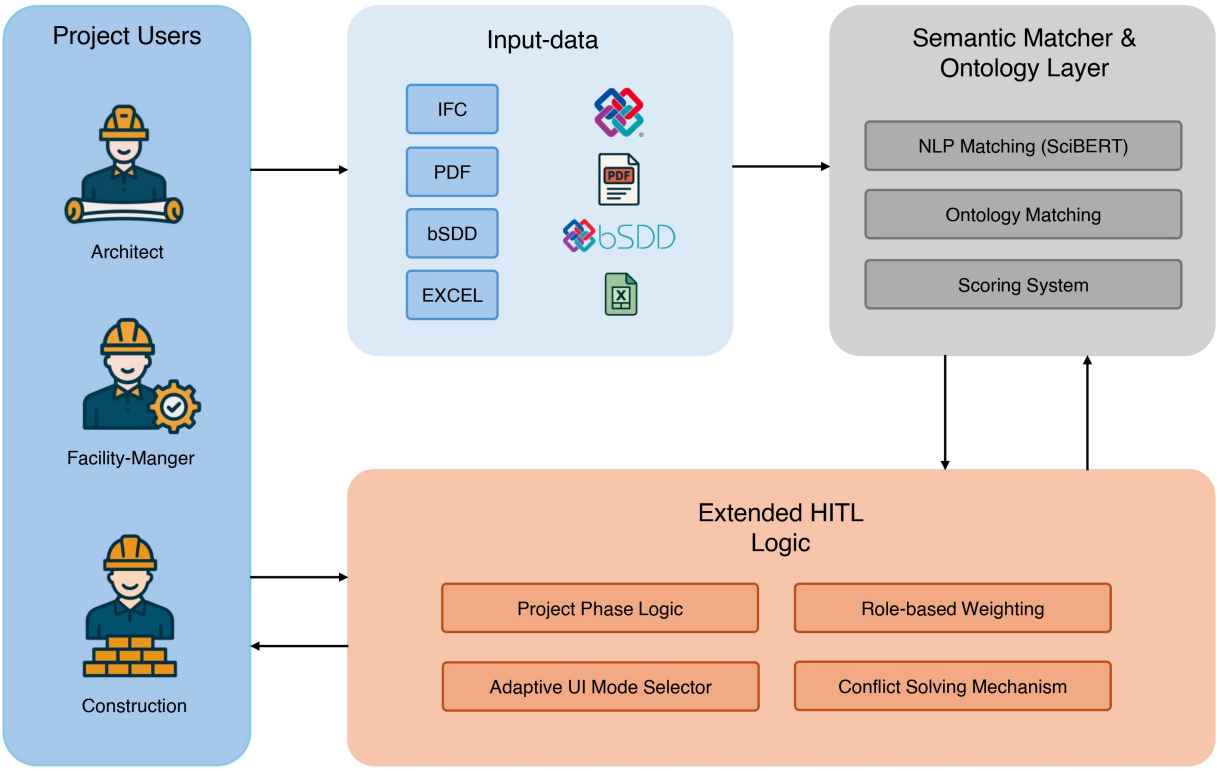


Figure 2: System architecture showing the interaction between project users, input data formats, the semantic matcher, and the HITL logic. The HITL module integrates feedback strategies tailored to user role, interface mode, conflict status, and project phase.

the system initiates second-level review, requests justification for feedback, or activates a consensus meeting mode to support collaborative resolution when necessary. All decisions are logged and used to inform future model behavior. These strategies are supported by a conflict resolution layer grounded in project hierarchies and informed by concepts from crowd annotation theory and escalation protocols, enabling structured and replicable disagreement handling.

Construction and FM projects evolve through distinct lifecycle phases with different data characteristics and stakeholder priorities. The system adjusts both algorithmic thresholds and role influence accordingly. In early design, broader matching is tolerated, and input from planners and designers is prioritized. During construction, execution feedback from roles like site managers gains importance in identifying practical constraints. In operations, stable datasets and performance monitoring elevate the role of facility management. Weight profiles are manually defined or inferred from metadata and blended during phase transitions. Similarly, the user interface and prompting strategy shift from exploratory guidance in the early stages to validation-focused input in the later stages.

4 Implementation

The extended HITL framework was integrated into an existing semantic matching platform for BIM and FM workflows [4]. The original system provided NLP-based similarity and ontology support; we added HITL components without changing the core.

The semantic matcher generates similarity scores using a combination of BERT embeddings, domain-specific patterns, and syntactic cues. These scores are processed by a scheduler that identifies low-confidence matches and generates review tasks. Tasks are routed to users by their role profiles stored in a central registry.

User feedback is stored together with the associated role information, enabling the system to prioritize corrections according to contextual authority. Submitted input incrementally updates similarity scores, category preferences, and ontology relations. All feedback actions are logged and visible to moderators. In the case of contradictory feedback, the affected match is marked as “in conflict” and displayed on a dedicated dashboard. Reviewers can then examine the feedback history, initiate clarification workflows, or issue binding decisions. If necessary, escalation rules notify higher-level roles for resolution.

The system also tracks the current project phase, which influences confidence thresholds and feedback interpretation. When phases change, new configurations are applied gradually to avoid abrupt changes in system behaviour.

To operationalize these mechanisms, the HITL module follows a recurring three-step decision cycle. First, model outputs are analysed and matches with low confidence or high uncertainty are flagged. Second, these cases are routed to users according to predefined role-based weighting factors, ensuring that the most relevant expertise is applied. Third, validated feedback is incorporated into the system through incremental weight updates, which in turn refine future suggestions. This cycle is implemented using modular scheduling components, making the process transparent and reproducible.

An event-driven architecture distributes feedback, updates, and conflicts, ensuring modularity. Prompts are context-aware and adapt to user roles: expert users can make detailed adjustments, whereas less experienced users receive guided suggestions. Tasks are prioritised by expected information gain and presented in small batches to avoid cognitive overload. All matching and feedback processes run asynchronously, allowing the system to continue processing while waiting for expert input.

5 Discussion

We focus on precision, recall, and error reduction; detailed metrics are in [4].

To assess the impact of the extended HITL framework, we compared it against a previously evaluated baseline system, which used a combination of pattern matching, lexical similarity, and a basic feedback mechanism. This baseline system was tested in a facility management scenario using 275 technical attributes across four domains (thermal, chemical, mechanical, geometric) and achieved an initial precision of 86.6 percent [4].

Table 1: Comparison of Baseline and HITL-Enhanced System Performance

Metric	Baseline System	With HITL Extension
Overall Precision	86.6%	91.8%
Overall Recall	81.4%	92.0%
Incorrect Matches	37 (13.4%)	26 (9.5%)
Avg. Confidence Score	0.79	0.87

Precision refers to the proportion of suggested matches that are correct, while recall measures the proportion of correct matches that the system successfully retrieves.

Building on the same dataset and architecture, the enhanced version incorporated role-weighted feedback, active learning, and phase-specific adaptation. As a result, the system achieved a recall of 92 percent and reduced incorrect matches in the final output by approximately 30 percent. These improvements were mainly driven by targeted human corrections and the resolution of conflicting feedback.

The active learning component played a central role by focusing user attention on uncertain cases. As the model integrated validated input over time, the number of required feedback queries decreased steadily, reducing user workload and improving efficiency. In contrast, the baseline HITL setup lacked prioritization and repeatedly prompted users for similar cases, resulting in a greater need for manual intervention.

Role-based feedback weighting also proved effective for resolving disagreements. In most cases, decisions aligned with high-authority input and matched expert ground truth. Escalation ensured corrections where needed.

Despite these positive results, several limitations and practical challenges remain.

Role-based weighting depends on accurate role attribution, which may be limited in heterogeneous or loosely structured environments.

Active learning and uncertainty estimation are well-established in text classification, but still emerging in construction informatics.

Overall, the evaluation demonstrates the potential of the proposed HITL framework to enhance semantic matching in complex construction workflows. At the same time, it highlights the need for continued refinement of role management, user interfaces, and feedback mechanisms to ensure robustness and usability across diverse project environments.

6 Conclusion and Future Work

This study shows that HITL enhances semantic matching in the AEC/FM Industry. By incorporating role-sensitive feedback, active learning, and phase-aware adaptation, the framework enables targeted expert input and improves both precision and robustness over purely automated methods.

Key contributions include a modular feedback architecture, context-aware interaction modes, and a conflict resolution mechanism grounded in domain authority. These components embed expertise without overwhelming users.

Future work will focus on several open challenges identified during evaluation. One priority is to improve the system's resilience in projects with inconsistent or missing role metadata, for example by exploring automatic role inference from behavior patterns or interaction history. Another direction is to enhance uncertainty estimation in sparse data domains through hybrid confidence metrics that combine semantic distance, data quality, and user correction history. Additionally, more adaptive user interfaces could improve accessibility, particularly for users with limited technical expertise.

7 Data availability statement

The data supporting this study's findings are available at https://gitlab.lrz.de/jmartin/HITL_extension_nlp_matching_tool, ensuring that they are Findable, Accessible, Interoperable, and Reusable (FAIR).

Acknowledgements

This research was supported by the Nemetschek Innovation Foundation and benefited from collaboration with industry partners within the framework of the openDBL project. The openDBL project is funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101092161, providing valuable test cases during development.

References

[1] P. Pauwels, "Supporting decision-making in the building life-cycle using linked building data", *Buildings*, vol. 4, no. 3, pp. 549–579, 2014. DOI: 10.3390/buildings4030549

[2] T. Bloch, "Connecting research on semantic enrichment of BIM – review of approaches, methods and possible applications", *Journal of Information Technology in Construction (ITcon)*, vol. 27, pp. 416–440, Apr. 2022. DOI: 10.36680/j.itcon.2022.020

[3] D. R. Honeycutt, M. Nourani, and E. D. Ragan, "Soliciting Human-in-the-Loop User Feedback for Interactive Machine Learning Reduces User Trust and Impressions of Model Accuracy", in *Proceedings of the Eighth AAAI Conference on Human Computation and Crowdsourcing (HCOMP-20)*, Virtual Conference: AAAI Press, 2020. [Online]. Available: https://ojs.aaai.org/index.php/HCOMP/article/view/7464

[4] J. Martin, C. Preidel, S. Vilgertshofer, and A. Borrmann, "Enhancing Construction Data Integration through Dynamic Ontology Alignment and Automated Attribute Mapping", in *Proceedings of the European Conference on Computing in Construction (EC³)*, Porto, Portugal: EC3 / CIB W78, Jul. 2025. [Online]. Available: www.dx.doi.org/%2010.35490/EC3.2025.250

[5] Z. Dragisic, D. O. Faria, C. Pesquita, J. Euzenat, and V. Svátek, "User validation in ontology alignment", *Journal of Web Semantics*, vol. 43, pp. 90–106, 2017. DOI: 10.1016/j.websem.2016.11.001

[6] M. Schröder, C. Jilek, and A. Dengel, "A Human-in-the-Loop Approach for Personal Knowledge Graph Construction from File Names", in *Proceedings of the Third International Workshop on Knowledge Graph Construction at the 19th Extended Semantic Web Conference (KGCW–ESWC 2022)*, ser. CEUR Workshop Proceedings, vol. 3141, Hersonissos, Greece: CEUR-WS.org, 2022. [Online]. Available: https://ceur-ws.org/Vol-3141/paper2.pdf

[7] J. Hussain et al., "Model-based adaptive user interface based on context and user experience evaluation", *Journal on Multimodal User Interfaces*, vol. 12, no. 1, pp. 1–16, Feb. 2018. DOI: 10.1007/s12193-018-0258-2 [Online]. Available: https://doi.org/10.1007/s12193-018-0258-2

[8] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, "Human-in-the-loop machine learning: a state of the art", *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3005–3054, Aug. 2022. DOI: 10.1007/s10462-022-10246-w