Semantic Interoperability of Heterogeneous Information Systems: A Hybrid Framework for Machine Learning-Assisted Semantic Mapping Recommendation

Aïcha KOULOU[1], Mohamed CHEKOUR[2], Norelislam EL HAMI[3], Nabil HMINA[4]
Higher School of Education and Training, Ibn Tofaïl University, Kenitra, Morocco[1, 2]
Science and Engineering Laboratory, ENSA, Ibn Tofaïl University, Kenitra, Morocco[3]
Ibnou Zohr University, Agadir, Morocco[4]

***Abstract*—Semantic interoperability is a major challenge in the integration of heterogeneous information systems, where differences in data structures, terminologies, and representations complicate the automatic identification of schema matches. Traditional schema matching approaches, primarily based on rules or lexical similarity measures, have limitations in the face of the increasing complexity of digital environments. This article proposes a hybrid framework to improve the automatic recommendation of semantic mappings by combining lexical, structural, statistical, semantic, and business similarities with a supervised learning model. The framework also incorporates expert validation to ensure the quality of the recommended matches and to feed an iterative process for improving the mapping repository. The evaluation is based on a case study in the health insurance sector. Experimental results show that the XGBoost model achieves the best performance with an F1 score of 89.48% and an AUC of 0.915. Top-k evaluations also highlight excellent recommendation capabilities, with a Hit@1 of 83.51%. The ablation study confirms that combining different similarity families significantly improves recommendation quality. These results demonstrate the value of the proposed framework for enhancing semantic interoperability, facilitating the identification of matches between heterogeneous schemas, and reducing the validation effort required from domain experts.**

***Keywords*—Semantic interoperability; schema matching; machine learning; XGBoost; information systems**

I. Introduction

Information needs to flow constantly across enterprise apps, cloud services, legacy systems, and decision-support systems. Even though they exist in the same organization, they are unlikely to be co-designed; they may originate from entirely different development lifecycles; they came potentially out of different software vendors and different units of the business, where based upon different data models, schemas, terminology, and business rules. Information exchange is a lot more complex than pure data transmission. On many occasions of system integration projects, the data could successfully get sent and exchanged via standardized interfaces or APIs, while it might imply different meanings in the other system.

On the other hand, identical-sounding concepts that may refer to identical items could be listed as separate items under different attribute names, structures, or coding schemes, leading to low data quality and consistency of business indicators, complicating automation workflow management systems; therefore, affecting the reliability of analytical and predictive applications. Information systems should be able to exchange and exploit information effectively, which is also known as interoperability [1]. Technical, organizational, legal, and semantic dimensions describe the interoperability issue [2]. Semantic interoperability could be the toughest to deal with, due to the fact of preserving the meaning of the exchanged information between heterogeneous systems.

One of the main requirements to achieve the objective of providing seamless data discovery across various content repositories is the ability of recognizing exact semantic matching across heterogeneous schemas, which has been investigated in the area of schema matching [3]. The schema matching techniques have been largely based upon human involvement, which results in expensive solutions based on manual specifications, whether as mappings, transformation rules, or reference dictionaries reflecting expertise in the domain or reference to the expert knowledge of humans. All of these have been efficient so far in addressing semantic matching of schemas of limited sizes, though the rapid growth in web services across various sectors is going to impose a heavy burden in maintaining these schemes through manual matching, as it costs money as well as time and effort. Also, such large-scale matching will make it even more difficult or impossible to rely on a single expert.

Recently, there has been progress in the field of machine learning, suggesting a different way of viewing this Semantic Matching Problem, that of a prediction problem. Instead of being only rule-based, the models can take lexical similarities, semantic similarities, structural similarities, statistical similarities, and business similarities to recommend the correspondences between databases to an expert that performs Semantic Matching.

A hybrid framework for interoperability-assisted by machine learning is proposed. Hybrid in that it attempts to integrate multiple similarity measuring techniques to be able to recommend matches between heterogeneous information

system fields while remaining under expert control through a human validation stage. To assess the reliability and practical relevance of the semantic mapping recommendations produced by the framework, the proposed approach is evaluated through a case study.

The remainder of this article is organized as follows. The next section presents related work on schema mapping, hybrid approaches, and entity mapping. This is followed by an introduction to the fundamental concepts of information systems interoperability, semantic interoperability, and data quality. The following sections will present the proposed framework and the results of an illustrative case study.

II. Related Work

Semantic interoperability has been studied through various research fields, such as schema matching, ontology alignment, and semantic integration.

As noted in [4], schema matching is a central problem in several domains, including data integration, data warehouses, e-commerce, and semantic query processing. Previous schema matching approaches primarily used lexical, syntactic, structural, or data type-based measures. The algorithm proposed by [5] represents a significant contribution to this category of works. It integrates various elements such as names, data types, constraints, and schema structure to establish mappings between heterogeneous components. This approach illustrates the benefit of not depending on one matching criterion, but rather integrating multiple relevant indicators to enhance the quality of the matches.

The system proposed by [6] supports various schema types, including relational and XML, and allows for the construction of more robust matching strategies from several elementary methods. What sets the hybrid system apart is its capability to counteract the drawbacks of isolated solutions. However, it usually remains dependent on manually defined parameters, combination rules, or thresholds, and its implementation in a real-world business context needs significant configuration, validation, and maintenance work.

The authors of [7] present ontology matching as a response to the problem of semantic heterogeneity encountered by computer systems. However, its use presupposes the existence of sufficiently comprehensive repositories that are maintained and adapted to the context under study. In operational environments, some concepts are highly organization-specific, which can limit the direct use of standard ontologies without business enrichment.

Machine Learning-based approaches have made it possible to reformulate matching problems as supervised learning tasks. Instead of manually defining all the matching rules, the model learns to distinguish matching pairs from non-matching pairs using annotated examples. The method proposed by [8] presents tools covering several stages of the process, including data preparation, blocking, matching, training, and evaluation. Similarly, the authors of [9] investigated the contribution of Deep Learning to entity matching, and more recently, [10] proposed using pre-trained Transformer-type language models for entity matching, reformulating the problem as a classification of text pairs. Recent research explores the use of language models to identify matches between elements of relational schemas based on field names and descriptions [11], [12], [13].

These contributions are important for this study because they confirm that matching problems can be treated as prediction tasks. However, they mainly focus on matching records or entities, while the framework studied here focuses on recommending correspondences between data fields and schema elements, in a logic of semantic interoperability.

III. Concepts

A. Information Systems Interoperability

Information systems interoperability can be understood as the capability for various systems or entities to exchange information and utilize it in a consistent manner. It is an essential condition for the functioning of modern organizations, particularly when business processes span multiple information systems or involve multiple actors. Interoperability frameworks emphasize the multidimensional nature of this concept. There are four levels of interoperability recognized by the European Interoperability Framework: technical, semantic, organizational, and legal [14]. The core of technical interoperability is protocols, interfaces, exchange formats, and communication tools. Organizational interoperability is defined as the synchronization of procedures, responsibilities, and goals among participant systems. Legal interoperability guarantees a trusted environment for exchanges (data protection, licenses, etc.). Semantic interoperability focuses on the meaning of the exchanged data and the ability of systems to assign the same meaning to it.

This study focuses on semantic interoperability because the technical connection between systems does not necessarily guarantee a correct understanding of the data. Two systems can exchange information even if the concepts they manipulate are not actually aligned. For example, two fields with similar names can refer to different realities, while two fields with different names can represent the same business concept. Semantic interoperability therefore aims to reduce these ambiguities by establishing reliable correspondences between the concepts and fields used by the systems.

B. Semantic Interoperability and Data Quality

Semantic interoperability is closely linked to data quality. When mappings between systems are poorly defined, the integrated data can become inconsistent, incomplete, or difficult to use. A poor mapping between two fields can lead to errors in automated processing, dashboards, statistical analyses, or predictive models. On the other hand, achieving accurate semantic alignment boosts data comprehension, simplifies data integration, and bolsters its adaptability for use in decision-making processes [15].

This issue aligns with the FAIR principles, designed to make data findable, accessible, interoperable, and reusable [16]. This set of principles underscores machines' capacity to automatically find, interpret, and use data. Seen from this angle, interoperability shouldn't be viewed merely as a technical transfer between systems, but also as a fundamental condition

for the intelligent use and reuse of data by analytical, decision-making, or predictive applications.

Semantic interoperability issues within organizations can manifest in several ways, including variations in field names, varied data formats, a lack of a shared reference system, differing coding values, divergent business terminologies, duplicate entries, or disparities in measurement units. These difficulties are common when several systems have been developed independently. The challenge then becomes building a mechanism capable of identifying semantic similarities between fields, while taking into account the business context in which these fields are used.

C. Schema Matching and Data Alignment

Schema matching is a central research area for data integration. It aims to identify correspondences between elements of heterogeneous schemas. These elements can be tables, columns, attributes, relationships, or entities.

Schema matching approaches can be classified according to several criteria. Schema-level matchers exploit information available in the schemas, such as names, types, or constraints. Instance-level matchers use observed values in the data. Element matchers compare isolated attributes, while structural matchers exploit relationships between multiple elements. Language-based matchers compare names and descriptions, while constraint-based matchers use types, keys, cardinalities, or rules associated with the data [4].

Hybrid approaches combine several types of signals to improve the quality of the matches [17]. This logic is particularly important because no single measure can handle all cases of heterogeneity.

IV. Proposed Framework

The proposed framework aims to support semantic interoperability between heterogeneous information systems through an automatic mapping recommendation mechanism between data fields. The goal is not to automatically generate definitive matches, but to provide a prioritized list of likely mappings. The system acts as a decision support tool for business, data, and IT teams. For each source field, it recommends the closest target fields based on several similarity dimensions. The business expert can then validate, correct, or reject the proposed matches. This approach aligns with human-in-the-loop methodologies, which combine the computational capabilities of models with human intervention in complex, ambiguous, or context-dependent tasks [18].

Therefore, the proposed framework is based on a hybrid approach. It combines metadata processing techniques, similarity measures, machine learning models, top-k recommendation, and human validation. This combination overcomes the limitations of purely manual approaches while preventing uncontrolled automation of semantic mapping.

A. General Architecture

The proposed framework is organized into nine main steps, as illustrated in Fig. 1. These steps make it possible to transform a manual alignment problem into a structured recommendation process.

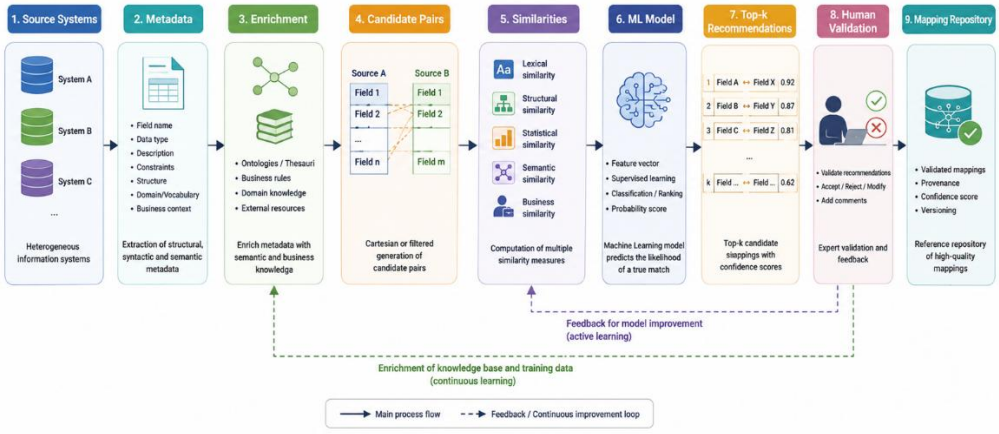

Fig. 1. Conceptual framework of the proposed approach.

B. Source Systems Mapping

The first step is to identify the information systems involved in the semantic alignment. This mapping helps define the scope of the study, the applications to be compared, the existing data flows, the business objects manipulated, and the repositories used.

The mapping is not limited to a technical inventory of the applications. It must also identify the business objects manipulated by each system.

The output of this step is a structured list of the systems, tables, entities, and functional domains involved in the alignment.

C. Metadata Extraction

After mapping, the framework involves extracting metadata associated with the data fields. This metadata forms the basis of the recommendation process. It may include:

- the field's technical name;
- the functional label;

- the business description;
- the data type;
- the table or entity to which it belongs;
- any constraints;
- the most frequent values or main categories;
- the rate of missing values;
- the number of distinct values;
- the associated functional domain.

Using metadata helps to limit reliance on raw values, which is particularly important when dealing with sensitive data.

D. Metadata Standardization and Enrichment

The extracted metadata must be normalized before comparison. This normalization process includes several steps: converting to lowercase, removing special characters, segmenting technical names, expanding abbreviations, translating if necessary, removing uninformative words, and harmonizing formats.

Enrichment adds supplementary information from industry dictionaries, synonyms, internal repositories, or sector-specific vocabularies. This step is important because raw lexical similarity does not always allow for the detection of semantic matches. Enrichment can also incorporate vector representations, such as text embeddings.

E. Generation of Candidate Pairs

If two systems contain n and m fields respectively, an exhaustive comparison produces $n \times m$ possible pairs. In a real-world environment, this number can become very high. The proposed framework therefore includes a candidate pair generation step to reduce the search space.

This step aims to retain plausible candidate pairs while filtering out pairs that are unlikely to represent valid semantic correspondences. Candidate filtering is performed using complementary criteria:

- data-type compatibility;
- functional-domain consistency;
- minimal lexical proximity;
- minimal semantic proximity;
- comparable business entities;
- presence of shared keywords, synonyms, or domain-specific terms.

A candidate pair is retained when it satisfies at least one of these compatibility conditions, thereby reducing the search space without excluding potentially valid mappings too early in the process. For example, a date field will be compared first to other date fields. A field belonging to the "contract" domain will be compared first to fields related to contracts, guarantees, or formulas. This filtering reduces unnecessary comparisons and focuses the model on truly plausible pairs.

F. Construction of Similarities

For each candidate pair, the framework constructs a feature vector representing several similarity dimensions. Let us consider two fields a and b, coming respectively from two systems S_1 and S_2. The candidate pair can be represented by:

$$P_{ab} = (\alpha,\beta) \quad (1)$$

This pair is associated with a vector of characteristics:

$$X_{ab} = [sim_{lex}, sim_{sem}, sim_{str}, sim_{stat}, sim_{met}] \quad (2)$$

where:

sim_{lex} represents lexical similarity;

sim_{sem} represents semantic similarity;

sim_{str} represents structural similarity;

sim_{stat} represents statistical similarity;

sim_{met} represents business similarity;

- Lexical similarity compares technical names and field labels. It can be calculated from measures such as edit distance, string similarity, token similarity, or word overlap.
- Semantic similarity compares the meaning of fields. It can be obtained from dictionaries, embeddings, or language models.
- Structural similarity takes into account the context of the field in the schema: table, entity, relationships with other fields, position in the data model, or proximity to certain business objects.
- Statistical similarity exploits the observed properties of the data, when these are available in raw or aggregated form: distribution of values, number of distinct values, average length, rate of missing values, dominant type or frequency of categories.
- Business similarity uses rules, standards, taxonomies, synonyms, or abbreviations specific to the field studied. It makes it possible to introduce functional knowledge into the learning process.

G. Machine Learning Scoring

Once the features are constructed, the problem is formulated as a supervised classification task. Each pair of fields is associated with a label:

$$y_{ab} = \begin{cases} 1 & \text{if } a \text{ and } b \text{ correspond to the same business concept} \\ 0 & \text{else} \end{cases} \quad (3)$$

The model learns a function:

$$f(X_{ab}) \rightarrow \Pi\,(y_{ab} = 1) \quad (4)$$

This function produces a matching score between 0 and 1. The higher the score, the more likely the pair is considered to be a match.

Several models can be used: logistic regression, Random Forest, XGBoost, SVM, or models based on textual representations. Random Forest models are capable of capturing nonlinear relationships between variables [19], whereas

XGBoost is generally used for tabular issues thanks to its efficient and scalable gradient boosting mechanism [20], [21]. Language models can also be utilized to boost semantic similarity or to generate textual representations of fields [10].

H. Recommendation Top-k

For each source field, the framework recommends the k target fields based on the highest scores. This top-k recommendation logic is suitable for operational use, as it enables the expert to select among several probable matches rather than relying on a single automatic suggestion.

Metrics like hit@k, and MRR are relevant for evaluating recommendation systems because they assess the ability of the system to place the correct match in the top results [11].

I. Human Validation and Improvement Loop

The proposed framework requires essential human validation to ensure the accuracy of mappings before their incorporation into operational processes. The framework recommends involving at least three domain experts, who assess recommendations using the canonical concept definitions, business rules, and domain dictionary as reference. When several experts are involved, agreement can be assessed using Fleiss' kappa, while unresolved cases can be addressed through consensus. Each recommendation can be assigned one of the following decisions:

- validated match;
- rejected match;
- match to be reviewed;
- partial or conditional match.

This validation reduces the risk of incorrect mappings, which can lead to integration or reporting errors. It also allows for the gradual accumulation of business knowledge. Validated decisions can be fed back into the training dataset to improve model performance in subsequent iterations.

The improvement loop can be represented as illustrated in Fig. 2:

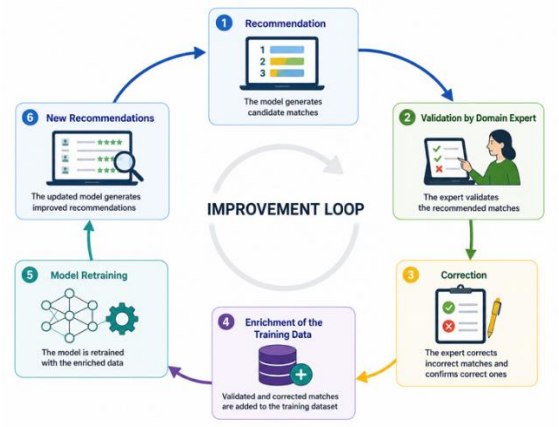

Fig. 2. Improvement loop for semantic mapping recommendation.

This loop allows semantic alignment to be embedded in a progressive, controlled, and traceable process.

J. Capitalizing on Validated Mappings

Validated mappings are stored in a mapping repository. This repository can be used in several IT applications:

- feeding ETL flows;
- harmonizing APIs;
- building a data dictionary;
- consolidating a business intelligence warehouse;
- populating dashboards;
- preparing analytical datasets;
- improving the reliability of predictive models.

Capitalizing on validated mappings avoids having to repeat the alignment process for each new integration project. It also contributes to improving data quality, understood as its suitability for use by data consumers.

V. Case Study and Results

In the illustrative case study, the proposed hybrid semantic interoperability framework assisted by machine learning was evaluated in a health insurance environment. This context is relevant because it involves several heterogeneous information systems that share common business concepts but use different structures, naming conventions, and repositories.

A. Data Used

The dataset used is structured around six systems (CRM, CONTRACT, CLAIMS, COMPLAINTS, DW, and PARTNER), 280 schema fields, 34560 candidate pairs, and 1250 positive mapping pairs. The resulting ratio is approximately 1:26.65, confirming the strong class imbalance inherent in the initial schema-matching dataset. The dataset used in the experiments was constructed from schema structures, terminology, and business knowledge representative of a confidential health-insurance environment. To preserve confidentiality, no raw personally identifiable production data were included. The 65 canonical concepts were defined in collaboration with domain experts and based on an internal business dictionary covering the main functional concepts shared across the participating information systems. Positive mappings were established by associating schema fields with these canonical concepts and validating the correspondences against their business definitions. Ambiguous cases were reviewed before being retained in the reference mapping set. Table I summarizes the main characteristics of the dataset.

TABLE I. Dataset Characteristics

Characteristic	Value
Information systems	6
Schema fields	280
Canonical concepts	65
Candidate pairs	34 560
Negative pairs	33 310
Positive pairs	1 250

To mitigate the strong class imbalance, controlled under-sampling was applied to the negative class. All 1,250 positive mappings were retained, while 2,500 negative pairs were selected, resulting in a ratio of 1:2.

To obtain a reliable evaluation and minimize information leakage, the dataset was partitioned at the canonical-concept level rather than by randomly splitting individual candidate pairs. Candidate pairs associated with the same canonical concept were assigned to the same subset, thereby reducing the possibility that highly similar instances could appear simultaneously in the training and evaluation sets. The resulting dataset was divided into training, validation, and test subsets using an approximate 70% /15% /15% distribution.

B. Experimental Implementation

All schema-field metadata were normalized by lowercasing, removing special characters, tokenizing technical identifiers, and expanding common abbreviations. Each candidate pair was represented by five similarity features normalized to [0,1]. Lexical similarity was computed using Jaro-Winkler similarity, while semantic similarity was obtained from cosine similarity between Sentence-BERT (all-MiniLM-L6-v2) embeddings. Structural similarity captured data-type compatibility and schema-context information, statistical similarity relied on differences in missing-value rate, cardinality, and average value length, and business similarity was derived from an internal health-insurance dictionary containing domain terminology, synonyms, and common abbreviations.

All supervised models were trained and evaluated using the same data partitions, feature representation, and class-balancing strategy. A fixed random seed of 42 was used for reproducibility. The main model configurations retained after validation are summarized in Table II.

TABLE II. MAIN MODEL CONFIGURATIONS

Model	Main hyperparameters
Logistic Regression	C=1.0, penalty='l2', solver='liblinear', max_iter=1000, random_state=42
Decision Tree	criterion='gini', max_depth=6, min_samples_split=10, min_samples_leaf=2, random_state=42
Random Forest	n_estimators=300, max_depth=8, min_samples_leaf=2, max_features='sqrt', random_state=42
Gradient Boosting	n_estimators=200, learning_rate=0.05, max_depth=3, subsample=0.8, random_state=42
XGBoost	n_estimators=200, max_depth=4, learning_rate=0.05, subsample=0.8, colsample_bytree=0.8, objective='binary:logistic', eval_metric='logloss', random_state=42

The binary decision threshold was also selected on the validation set by maximizing the F1-score and was subsequently fixed for test evaluation. For Top-k recommendation, candidate mappings were ranked directly according to their predicted probabilities.

C. Performance of Classification Models

To compare different match detection strategies, several models were evaluated. The first baseline model relied solely on lexical similarity. The second baseline model calculated an average hybrid score based on the various available similarities. Supervised models were then tested: logistic regression, decision tree, random forest, gradient boosting, and XGBoost.

Table III presents the results obtained on the test set.

TABLE III. RESULTS OF THE PERFORMANCES OBTAINED

Model	Accuracy	Precision	Recall	F1-score	AUC
Lexical similarity	82.47 %	72.11 %	69.68 %	70.87 %	0.748
Hybrid scoring (rules)	86.53 %	80.74 %	81.91 %	81.32 %	0.836
Logistic regression	88.28 %	84.36 %	84.04 %	84.20 %	0.861
Decision tree	87.71 %	82.64 %	83.51 %	83.07 %	0.842
Random Forest	90.76 %	86.87 %	87.23 %	87.05 %	0.889
Gradient Boosting	91.85 %	87.94 %	89.36 %	88.64 %	0.903
XGBoost	92.90 %	88.54 %	90.43 %	89.48 %	0.915

The results show that lexical similarity used in isolation provides limited performance. Although this approach achieves an accuracy of 82.47%, it remains heavily influenced by class imbalance. In contrast, metrics better suited to this context, such as precision (72.11%), recall (69.68%), and F1 score (70.87%), highlight the difficulty of purely lexical approaches in correctly identifying semantic matches.

Integrating several families of similarities significantly improves performance. The hybrid score, which combines lexical, structural, statistical, semantic, and domain similarities, achieves an F1 score of 81.32%, an improvement of more than 10 percentage points over the lexical approach.

Supervised models confirm this trend by achieving superior performance. The XGBoost model achieves the highest performance with an accuracy of 92.90%, a precision of 88.54%, a recall of 90.43%, an F1-score of 89.48%, and an area under the ROC curve (AUC) of 0.915. These results reflect an excellent ability of the model to distinguish correct matches from non-matches, even in a context characterized by a strong imbalance between classes.

D. Results of the Top-k Recommendation

Beyond binary classification, the proposed framework aims to recommend a prioritized list of potential mappings. Top-k metrics were calculated on source field groups for which at least one positive match exists in the test set. The results are presented in Table IV:

TABLE IV. TOP-K RESULTS

Model	Hit@1 (%)	Hit@3 (%)	Hit@5 (%)	MRR
Lexical Similarity	60.85	73.40	78.72	0.663
Hybrid Similarity	73.19	83.51	86.70	0.773
Logistic Regression	75.53	84.57	87.77	0.791
Decision Tree	71.81	81.91	85.11	0.756
Random Forest	79.79	87.77	89.89	0.835
Gradient Boosting	81.91	89.36	91.49	0.852
XGBoost	83.51	90.43	91.49	0.871

The Top-k evaluation further demonstrates the effectiveness of the proposed semantic mapping approach. The lexical baseline achieves a Hit@1 of 60.85%, indicating that the correct mapping is ranked first in only about three-fifths of the queries, while the Hit@3 (73.40%) and Hit@5 (78.72%) scores suggest that relevant mappings are often retrieved but not consistently ranked at the top. Combining lexical, structural, statistical, semantic, and domain-specific similarities substantially improves the ranking quality, with the hybrid similarity score increasing Hit@1 to 73.19% and achieving Hit@3 and Hit@5 scores of 83.51% and 86.70%, respectively. Supervised learning models further enhance the ranking performance, with ensemble methods outperforming both lexical and hybrid approaches. Among them, XGBoost achieves the best overall results, reaching a Hit@1 of 83.51%, a Hit@3 of 90.43%, a Hit@5 of 91.49%, and an MRR of 0.871, demonstrating its ability to consistently rank the correct semantic correspondence among the highest-ranked candidates and thereby reduce the effort required for manual validation.

E. Similarity Ablation Analysis

To assess the contribution of each family of similarities, an ablation analysis was performed using the XGBoost model, identified as the most effective during the comparative evaluation (see Table V). Each variant consists of progressively adding a new family of features while maintaining the same hyperparameters and the same experimental protocol.

TABLE V. ABLATION ANALYSIS

Variant	Similarities used	Precision	Recall	F1-score	AUC
V1	Lexical	72.11%	69.68 %	70.87 %	0.748
V2	Lexical + Structural	78.62 %	77.13 %	77.87 %	0.815
V3	V2 + Statistical	83.15 %	84.04 %	83.59 %	0.861
V4	V3 + Semantic	86.91 %	87.77 %	87.34 %	0.897
V5	Complete Framework (Lexical + Structural + Statistical + Semantic + Business)	88.54 %	90.43 %	89.48 %	0.915

The ablation results show a progressive improvement in performance as additional similarity families are integrated. Using lexical similarity alone yields the lowest F1-score (70.87%). Adding structural similarity produces the largest incremental improvement, increasing the F1-score by 7.00 percentage points to 77.87%, while the addition of statistical similarity provides a further gain of 5.72 points. Semantic similarity then increases the F1-score to 87.34%, and the final inclusion of business similarity further improves it to 89.48%. These results indicate that all similarity families provide complementary information, with structural and statistical similarities contributing the largest incremental gains, while semantic and business similarities provide additional improvements that lead to the best overall performance of the complete framework.

The results illustrate the approach of developing a hybrid framework for solving the problem of semantic interoperability using heterogeneous information systems. Based on the evaluation metrics, lexical, structural, statistical, and semantic/business similarities combine together, resulting in the high accuracy of matching schema correspondences as opposed to the solution with single features together. As a result, cross-comparison analyses and top-k recommendation, as well as ablation study results, validate the relevance of this method with emphasis drawn on business knowledge and semantic similarities. These findings thus confirm the framework's capability to enable the identification and validation of semantic alignments, marking a crucial advancement towards enhanced semantic compatibility between information systems.

VI. DISCUSSION

The results obtained confirm that improving semantic interoperability cannot rely solely on lexical comparison of schemas. Differences in terminology, structure, and business context between information systems necessitate an approach capable of leveraging complementary information sources. In this regard, the proposed hybrid framework demonstrates that a combination of lexical, structural, statistical, semantic, and business similarities enables the generation of more relevant and robust recommendations for identifying correspondences.

In addition to the performance attained, this investigation underscores the benefit of a decision-making approach supported by machine learning. The collaboration between artificial intelligence and business expertise represents a significant lever for accelerating data integration projects while improving the quality of semantic mappings.

The evaluation was conducted within a health-insurance context, which may limit the direct generalization of the results to other application domains. Although the proposed framework is domain-independent in its overall architecture, some components, particularly the business similarity and domain dictionary, require adaptation to the terminology and business rules of each sector. Future evaluations should therefore consider additional domains such as logistics, financial services, or healthcare information systems to further assess the generalizability of the approach.

Future work may also focus on taking into account the evolution of business patterns and vocabularies, in order to maintain a high level of semantic interoperability in constantly evolving information systems.

VII. CONCLUSION

This study presented a hybrid semantic mapping recommendation framework aimed at improving semantic interoperability between heterogeneous information systems. The proposed approach integrates multiple types of similarities (lexical, structural, statistical, semantic, and business) with a supervised learning model to automate mapping recommendations, while keeping the business expert at the heart of the validation process.

Experimental evaluation, conducted on a case study in the health insurance sector, highlights the effectiveness of the proposed framework. The results show that combining complementary information sources significantly improves the quality of recommendations compared to approaches based on a single similarity measure. Comparative analyses, Top-k recommendation performance, and ablation studies confirm the contribution of each family of features to the quality of the recommended mappings.

Beyond the performance achieved, this research underscores the value of a hybrid approach for facilitating data integration projects. By proposing relevant semantic correspondences and reducing the effort required for their validation, the developed framework provides decision support that can accelerate semantic interoperability processes while improving their reliability.

REFERENCES

[1] G. Lalli, "Defining Interoperability: A Universal Standard," in Journal of ICT Standardization, vol. 13, no. 2, pp. 139-156, June 2025, doi: 10.13052/jicts2245-800X.1323.

[2] R. S. P. Maciel, P. H. D. Valle, K. S. Santos, and E. Y. Nakagawa, "Systems interoperability types: A tertiary study," ACM Computing Surveys, vol. 56, no. 10, pp. 1–37, 2024, doi: 10.1145/3659098.

[3] Norazian M Hamdan and Novia Admodisastro. "Towards a Reference Architecture for Semantic Interoperability in Multi-Cloud Platforms". International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 14, No. 12, 2023.

[4] E. Rahm and P. A. Bernstein, "A survey of approaches to automatic schema matching," The VLDB Journal, vol. 10, no. 4, pp. 334–350, 2001.

[5] J. Madhavan, P. A. Bernstein, and E. Rahm, "Generic schema matching with Cupid," in Proceedings of the 27th International Conference on Very Large Data Bases, 2001, pp. 49–58.

[6] H.-H. Do and E. Rahm, "COMA: A system for flexible combination of schema matching approaches," in Proceedings of the 28th International Conference on Very Large Data Bases, 2002.

[7] J. Portisch, M. Hladik, and H. Paulheim, "Background knowledge in ontology matching: A survey," Semantic Web, vol. 15, no. 6, pp. 2639–2693, 2024, doi: 10.3233/SW-223085.

[8] P. Konda et al., "Magellan: Toward building entity matching management systems," Proceedings of the VLDB Endowment, vol. 9, no. 12, pp. 1197–1208, 2016.

[9] S. Mudgal et al., "Deep learning for entity matching: A design space exploration," in Proceedings of the 2018 International Conference on Management of Data, 2018.

[10] Y. Li, J. Li, Y. Suhara, A. Doan, and W.-C. Tan, "Deep entity matching with pre-trained language models," Proceedings of the VLDB Endowment, vol. 14, no. 1, pp. 50–60, 2020.

[11] E. Sheetrit, M. Brief, M. Mishaeli, and O. Elisha, "ReMatch: Retrieval enhanced schema matching with LLMs," arXiv:2403.01567, 2024.

[12] M. Parciak et al., "Schema matching with large language models: An experimental study," arXiv:2407.11852, 2024.

[13] O. E. Gungor, D. Paulsen, and W. Kang, "SCHEMORA: Schema matching via multi-stage recommendation and metadata enrichment using off-the-shelf LLMs," arXiv:2507.14376, 2025.

[14] L. Daou, E. Petitdemange, G. Zacharewicz, N. Daclin, and S. Durieux, "Structuring federated data interoperability: A multi-level framework," Procedia Computer Science, vol. 274, pp. 664–675, 2025, doi: 10.1016/j.procs.2025.12.065

[15] B. H. de Mello, S. J. Rigo, C. A. da Costa, R. da R. Righi, B. Donida, M. R. Bez, and L. C. Schuh, "Semantic interoperability in health records standards: a systematic literature review," Health and Technology, vol. 12, no. 2, pp. 1–16, Mar. 2022. doi: 10.1007/s12553-022-00639-w.

[16] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, vol. 3, article 160018, 2016.

[17] A. Koulou, M. Zemzami, N. E. Hami, A. Elmir, and N. Hmina, "Optimization in collaborative information systems for an enhanced interoperability network," Int. J. Simul. Multidisci. Des. Optim., vol. 11, p. 2, 2020.

[18] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," Future Generation Computer Systems, vol. 135, pp. 364–381, 2022.

[19] H. A. Salman, A. Kalakech, and A. Steiti , Trans., "Random Forest Algorithm Overview", Babylonian Journal of Machine Learning, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.

[20] Y. Koulou and N. El Hami, "Machine learning for recommender systems under implicit feedback and class imbalance," International Journal of Advanced Computer Science and Applications, vol. 16, no. 9, 2025, doi: 10.14569/IJACSA.2025.0160954.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.