Authors' responses to the feedback of the reviewers
Paper: From Manual to LLM-assisted: A Mapping Review on Ontology Construction Methods

First of all, we would like to sincerely thank the Editor and the reviewers for their time, effort, and careful reading of our manuscript.

We appreciate the constructive comments and suggestions provided by the reviewers. We have carefully considered all remarks and revised the manuscript accordingly.

Please find detailed point-by-point responses below.

Reviewer 1

Comment 1. The clustering of ∼41 manual methods into k=11 clusters produces many singletons and pairs whose practical value is unclear. The conclusion is very brief and offers no synthesis across categories, no identification of open problems, and no discussion of limitations.

Response: We thank the reviewer who gave us the opportunity to further detail this aspect in the paper. Indeed, we agree that k = 17 (11 was a typo in the text; in the images, the value was correct) is not that informative *per se*. There are other k values that, although reaching a lower score, avoid the creation of singletons. What we aimed to represent there is that the best k (but not only) is capable of clustering similar construction methodologies. For visualisation as well, we did not consider it profitable to show pictures of several clustering varying the k value, also because choosing to visualise the cluster of another k that is not based on the silhouette score would make our choice questionable or subjective. Hence, we took the chance in the Discussion of that section to clarify that the role of the visualisation with k = 17 is to demonstrate that the methodology is able to cluster similar methodologies, although maybe for some users, other splits may be more useful. We also introduced a global Discussion at the end.

Comment 2. No data deposit or long-term URL is provided, which the SWJ survey track explicitly requires.

Response: We are happy to introduce a URL to allow other researchers access to the resource: https://github.com/davide1797/ontology_constructions. We also added it to the manuscript.

Comment 3. Some figures (especially 10 and 11) are difficult to read at print size, and a copy-editing pass is needed for minor errors throughout.

Response: We took the opportunity to enlarge the font size and points of Figures 10 and 11. We hope now they are clearer. We also fixed small grammar errors and typos.

Comment 1. With k=11 clusters and several singletons, what is the practical guidance a reader should draw from the clustering?

Response: We thank the reviewer for this relevant observation. We agree that the practical interpretation of the clustering results was not sufficiently explained in the previous version of the manuscript. The objective of the clustering analysis was not to propose a rigid taxonomy of ontology construction methodologies, but rather to provide practical guidance by identifying groups of approaches sharing similar characteristics. In the revised manuscript, we clarified that large clusters represent recurrent methodological patterns and can help practitioners select approaches according to their needs. Singleton clusters should not be considered as clustering errors; instead, they correspond to methodologies presenting distinctive features or original design choices compared to the rest of the literature. We have added a discussion paragraph after the clustering results to better explain the interpretation and practical value of these clusters. In a nutshell, the best k is used to visualise an example of clustering of methodologies rather than information by itself, and that singletons are created for peculiar characteristics of some methodologies. With different k values, we have not encountered singletons, but we chose k = 17 anyway since it resulted in the best k in a more objective and numerical interpretation.

Comment 2. Why was Gower's distance not considered for the mixed categorical/binary/ordinal feature space instead of the custom distance function?

Response: We recognise that our distance function was not standard and we are grateful to the reviewer for this suggestion. We realised, indeed, that the proposed distance is very similar to the one proposed by us, but with the introduction of the normalisation that can make the learning process smoother for categories with many labels. In our case, almost all the features were binary. We repeated the experiments with the proposed metric, and we updated the manuscript accordingly, with 18 now being the best k for the clustering.

Comment 3. For the semi-automatic LLM-based classification, was any robustness check performed across different LLMs or prompts?

Response: We thank the reviewer who gave us the opportunity to better detail that part. Indeed, we showed the result of one model, but we also tested other ones like Sonnet and Gemini. The most interesting result was that they all agreed on 4 clusters, and the categories were sometimes very similar or overlapping. We reported this result and our query on the paper for completeness and reproducibility, and the various answers in a dedicated folder in the repository. We also invite the Reviewer to have a look at the response to the Reviewer 2 Comment 4.

Comment 4. Could you provide a data deposit containing encoded feature tables, clustering scripts, LLM prompts/responses, and README?

Response: The requested material is now available following the GitHub link: `https://github.com/davide1797/ontology_constructions`.

Comment 5. Could the conclusion be expanded to discuss cross-category findings, concrete open problems, and limitations?

Response: We introduced an inter-category evaluation and exploit it also to evaluate findings, problems, limitations, and the potential utility of this contribution in the future.

Comment 6. The representation of clusters in Figure 8 should be improved.

Response: We took the opportunity to regenerate the image based on the new classification and to improve its readability.

We thank Reviewer 1 for the constructive and encouraging comments, especially for helping us improve the clustering process and for suggesting that we demonstrate robustness in the LLM-based classification.

Reviewer 2

Comment 1. Potentially out-of target. A survey must be introductory to a certain extent. However, in this case (Sections Introduction and Background Knowledge) the paper appears to be too introductory in defining what is an ontology and a KG. In contrast, a contextualization and scope of the survey at a general level is missing.

Response: We thank the Reviewer for this remark. We indeed agree that the introductory part did not provide sufficient context, information, and scope of the survey, postponing it to the next sections. We took the opportunity to clarify some aspects that are, in our opinion, better detailed in the remainder of the manuscript.

Comment 2. Engagement with existing literature. The NLP literature presented in the introduction (e.g. LeCun et al, 2015) feels very broad and somehow generic. I expected to find some pointers to existing challenges and approaches at the intersection between SW and NLP to be fully examined in the other sections. Additionally, existing surveys that overlap with this one are not presented. For instance:
- Zengeya, T., & Fonou-Dombeu, J. V. (2024). A review of state of the art deep learning models for ontology construction. IEEE Access, 12, 82354-82383

- Li, J., Garijo, D., & Poveda-Villalón, M. (2025). Large language models for ontology engineering: a systematic literature review. Semantic Web Journal, 20(x), 1-45

Response: We thank the Reviewer for this remark. We took the chance to broaden the context in the NLP field, which is evidently one of the most impacted domains by these advancements. We also thank the Reviewer for the two suggested papers. We actually were aware of the second one, but we waited for its acceptance before introducing it in the paper. They have been discussed in the Introduction in their complementary aspects with respect to our proposal.

Comment 3. Scattered methodology. The presentation of the survey method is clear. However, in the following sections a number of additional methodological approaches are adopted to analyse papers without a specific rationale. For instance, clustering methods at page 9. I suggest to create a methodology section that can include all the methodologies adopted in the survey, being clearer about the authors' design choices.

Response: We thank the Reviewer for this remark. We agree that the intra-category classification approaches and choices were not explicit or clarified. After the picture of the general methodology, we deepened the intra-category classification by enriching the paragraph *Selection & Evaluation* on page 6 and with another specific schema.

Comment 4. Problematic usage of LLM-as-judge. In two different sections (Review Methodology and Semi Automatic Ontology Creation) authors introduce the usage of a LLM (GPT) to automatize their analysis. This is not necessarily problematic. However, if authors propose a LLM-as-judge approach they should ensure that they properly checked the model's output and mitigated potential bias in the pipeline. Additionally, using a commercial closed-sourced model makes the methodology non replicable.

Response (same as Reviewer 1 Comment 3): We thank the reviewer who gave us the opportunity to better detail that part. Indeed, we showed the result of one model, but we also tested other models sucha as Sonnet and Gemini. The most interesting result was that they all agreed on 4 clusters, and the categories were sometimes very similar or overlapping. We reported this result and our query on the paper for completeness and reproducibility, and the various answers in a dedicated folder in the repository.
(in addition): We also accept the suggestion of proposing alternative free models to perform the same evaluation. In the manuscript, we left the result that we considered the best one in terms of label clarity and mapping, but we showed some consideration based on the repetition of the experiment and some Sankey diagrams to visualise the correspondences of the clusters of the different models. In the repository, we added all the results provided by the models.

Comment 5. Scope of the survey. While the general aim of the survey is clear, its scope is not entirely understandable. Specifically, RQs are presented without a specific connection to the objectives of the manuscript. This hinders following the flow of the narrative.

For instance, "RQ6: What is the context of the methodology?" is not explained nor logically connected to the overall work leaving me wondering why this question was relevant for the survey. Dedicating more space to contextualize the selection of these RQs is essential to improve the manuscript making its contribution clearer and easier to follow.

Response: We better clarified the rationale behind the decision of our RQs, explaining that they represent to us the more relevant questions when choosing a methodology, with the last one to distinguish it with respect to the others. We made this clearer by making a reference between objectives and RQs.

Comment 6. Thoroughness of discussion. The three sections where results are presented are very descriptive. They briefly answer the questions with lists of items (e.g., type of evaluation approaches) without discussing them. For instance, in the Semi-Automatic Section (page 9), "RQ4: What metrics are available?" abruptly refers to Table 7 without further discussion about these metrics. In other cases (Figure 9, right, page 13) the survey appears to rely on some taxonomical shortcuts creating the category "Specific procedure" that might mean a wide range of things. I suggest rethinking all these sections trying to identify some takeaways that one can get from the survey. E.g., Table 12 (page 14) on automatic methods for ontology construction shows a clear imbalance towards reference-based metrics (F1, precision, recall, accuracy), which in the actual AI landscape are contested since they are not reliable to assess models' abilities out-of-domain and their biases (e.g., a LLM can generate an ontology that is biased towards specific socio-demographics). This is an interesting insight about some directions that could be taken to design better evaluation metrics.

Response: We thank the Reviewer for this comment; indeed, we now provide more meaningful insight and information for every question instead of just referring to the table. This allowed us to summarise some aspects and pose some open questions. We also agree that "Specific procedure" was not that informative; we meant some specific mapping based on the data structures of their dataset. We renamed the label "Data Mapping".

Comment 7. Some colloquial passages. The paper is clearly written but I suggest revising some passages that are too colloquial. E.g. "It is not difficult to imagine that many papers are in common between the different sources"; "as you can see from Figure 8".

Response: We thank and agree with the Reviewer for this remark. We reformulated some colloquial passages and fixed small typos and grammar errors.

Comment 8. No permanent links provided. The paper presents several automatic methods supporting the survey but no link to the code is provided nor a pointer to the list of selected papers.

Response (same as Reviewer 1 Comment 2): We have now introduced a URL to allow other researchers access to the resource: `https://github.com/davide1797/ontology_constructions`. We also added it to the manuscript.

We thank Reviewer 2 for his comments, especially to make us improve our manuscript in better clarifying the context, positioning, the evaluation methodology and the taken decisions, which evidently impacts the overall presentation.

Reviewer 3

Comment 1. Most of the research questions are answered through descriptive reporting of the observed data rather than through insights derived from the extracted and analysed evidence.

Response (same as Reviewer 2 comment 6): We thank the Reviewer for this comment; indeed, we now provide more meaningful insight and information for every question instead of just referring to the table. This helped us to summarise some aspects and pose some open questions.

Comment 2. The exact criteria used to exclude papers during the screening phase are unclear. Additionally, it is not explained how these criteria were applied to papers retrieved from arXiv.

Response: We thank the Reviewer for this remark, which allows us to clarify this aspect. The pipeline of paper selection (pp. 4-6) proposes three steps after the *Database Lookup* that returned about 5 000 papers:

- *Screening*: we asked an LLM to filter out irrelevant papers for the scope of the survey. We manually checked and, in some cases, reintegrated excluded papers. The result was less than 1 000 papers.
- *Duplicates Removal*: with a script verifying titles, first author, and year, we automatically removed almost all the duplicates. In this phase, we also manually ruled out papers that were explicit extensions that did not introduce novelty for the scope of the survey. At this point, we had about 400 papers.
- *Inherence Filtering*: by looking at the abstract, title, conclusions and, if necessary, the entire content, we manually filtered the papers of our interest for the survey. We ended up with 136 papers. We agree that maybe this part was poorly described, also given the subjective nature of the selection. We clarified the paragraph with more details and criteria of how this has been performed.

As the comment suggests, we did not distinguish the venue of the selected papers; hence, no particular conferences or journals were privileged, and arXiv papers have been analysed as the other ones. We claim this choice was needed, especially for recent works on LLMs, in which several proposals exist but may not be mature enough.

Comment 3. Did the authors apply snowballing as part of the search strategy?

Response: For the generic search of papers, we did not apply snowballing, but we used it when we fetched reviews. In that sense, we looked for cited papers not returned by our query. However, the number of additional retrieved papers was negligible with respect to the total number of papers, so we did not specify the number of works introduced by snowballing in the review. We wrote this as part of our strategy, but we took the opportunity to better clarify it in the *Inherence Filtering* paragraph on page 6.

Comment 4. In "Step 8. Local Classification", the parameters and metrics used for comparison should be explicitly listed and defined.

Response (same as Reviewer 2 comment 3): We thank the Reviewer for this remark. We agree that the intra-category classification approaches and choices were not explicit or clarified. After the picture of the general methodology, we deepened the intra-category classification by enriching the paragraph *Selection & Evaluation* on page 6 and with another specific schema.
(in addition): we claim that the three intra-category classifications need specific classification criteria; hence, no generic metrics for the three are proposed. Conversely, we specified 8 features for the first one, listed and described on page 8, and the criteria to classify the other two categories, respectively, on pages 12 and 18 (RQ_8).

Comment 5. The paper provides numerous tables and figures; however, it is not clear how these artifacts were generated or how they can be traced back to the analyzed papers. All extracted data features should be described, together with the analysis and harmonization procedures applied. For example, what observations from the primary studies led to the generation of Table 2? The same concern applies to the tables and figures associated with Figures from 6 on. The data collection and validation procedures should also be explained.

Response: We thank the Reviewer for this remark. We took the opportunity to clarify, for each category, the extracted information. For Table 2, also to answer another comment, we provided more information about the collecting process and information to fill it. For the first category, we searched in the selected papers the selected F1 - F8 features. The description of the process allowed us to fill these features with their proper values (Table 6). This has been done by seeking explicit and implicit information in the papers. For instance, a methodology may explicitly state that it requires the existence of knowledge (documents) or it can be implicit in some description of the steps. The definition of every feature is available on page 8 (RQ_8). For the second category, we sought the strategies adopted for the manual and automatic parts. We proposed some common labels to distinguish them, and we used them for the subsequent classification steps. We agree that a clarification of the labels was not present here; so, we added this clarification and definitions of the labels on page 12 (RQ_8), which also helps understand the harmonisation of the labels. For the automatic category, we sought the solved tasks and the adopted technologies to solve them, and the classification is based on these two dimensions (Figures 13 and 14). Here, the labels are extracted from extracted reviews or directly from the cited paper. We better clarify the information search in every RQ_8 paragraph and also added in the Appendix the created dataset for the semi-automatic part, which indeed was not present in the first version. We took the opportunity throughout to trace every result with the corresponding data by providing a link to some tables or to our repository.

Comment 6. It should be clearly stated which data were extracted from the papers (and their corresponding values) to answer each research question.

Response: The extracted data correspond to the listed research questions on page 4. By reading the papers, we analysed the aspects of interest, like limitations, evaluations, metrics, context, and so on. For the visualisation, we preferred a summary table for every research question within each category for the RQ_4 - RQ_7. For the classification (RQ_8), depending on the category, we proposed a different classification and, according to the case, proposed a visualisation to correlate the selected papers. We enriched the *Selection & Evaluation* paragraph on page 6 to better highlight this aspect.

Comment 7. How is the term "feature" defined in Table 2?

Response: Table 2 shows a list of reviews on manual ontology construction. The intended goal is to show the classification features of those papers, some of which are reused in our manuscript to classify papers in that category. We clarified this aspect in the text (page 7, RQ_4) and by changing the title columns of Table 2.

Comment 8. Another major issue concerns the predefined categories of manual, semi-automatic, and automatic approaches. Are all retrieved works complete methodologies, or is there a mix of methodologies and methods/techniques targeting specific tasks or activities? If so, are all these elements truly comparable?

Response: We thank the Reviewer for this interesting point. For this, we recall that the review only includes methodologies for ontology construction or learning. If a methodology comprises both manual and automatic parts, it clearly fits the definition of a semi-automatic approach, while the mix of manual (resp. automatic) parts would lead to a bigger manual (resp. automatic) approach. We also emphasise that the distinction of the three categories is justified in literature by the presence of existing works explicitly clarifying the context, like "*Manually vs semiautomatic domain specific ontology building*", "*A systematic literature review of automatic ontology construction*", or "*Semi-automatic ontology construction based on patterns*". We claim this has been necessary in the literature exactly to restrict the context of comparable works, given the difficulty of comparing methodologies belonging to different categories. This is also reflected by the three different evaluations provided in the manuscript. We consider this reflection worth mentioning in the manuscript in the added *Inter-category Discussion* section.

Comment 9. Why is "linguistic" considered a category within Type (F1)? How was the list of features defined?

Response: The list of features (F1 - F8) on page 8 has been chosen by taking inspiration from existing features of the cited review papers, and by introducing some other features of interest in our opinion, like *agility*. The goal is to propose a set of features that help ontology designers choose a proper methodology based on their context, skills, or application. Strictly speaking of F1, the goal of this metric is to clarify a macro-objective of the methodology. We easily identified methodologies being developed for a specific context or application (project), others aiming to be fully reusable (framework), and others tailored to reuse existing sources (integration). When reading about ontology construction in linguistics, we realised that none of these types fitted it. Building linguistic ontologies is neither limited to an application nor aiming to be a general framework. It is, generally, a construction intended to integrate existing linguistic resources, but it is not only that: it also needs to understand which (linguistic) tasks are to be executed, how to align semantics (many versions of synsets are available), the size of the integration, and so on. This is why we claim that reducing linguistic construction to integration would be too simplistic and that it deserves its own value in that feature. We thank the Reviewer for this remark, which helped us in explicating this choice in the paper on page 8 (RQ_8).

Comment 10. It is not clear what is meant by "local evaluations."

Response: We agree with the Reviewer: we actually used two terms to refer to the same process. We aligned "local evaluations" with what we called "intra-category classification". With this term, we mean the classification of the papers within the categories in which they have been put.

Comment 11. Page 9: "For this, we reduce the 8 features to 3." Why was this reduction performed, and which features were retained?

Response: We adopted a robust technique in the literature for dimensionality reduction called Principal Component Analysis (PCA). The advantage of this technique is that we did not need to choose 3 features among the 8, but rather it synthesises the information of the original 8 features into 3 (hidden) features. There is, naturally, a loss of information in performing this process, but in the case of selecting 3 features out of 8, this is negligible. Moreover, this process has been done only for visualisation purposes, in order to show Figure 7 (b).

Comment 12. In general, the presentation of partial information in tables creates uncertainty regarding whether all papers were analyzed. For example, should readers assume that papers not appearing in Table 5 have no reported limitations, or that the table only presents a subset of the results? Information should be reported for all studies, including the explicit absence of particular characteristics.

Response: We are actually considering this matter. The tables represent complete information: missing papers simply do not contain any informative part to answer the specific RQ. For instance, in the context table, non-cited papers are works for which a context has not been specified. We agree, however, that tables may be interpreted ambiguously. The main concern is that, for some tables, the papers for which the information is missing represent the majority, and much space would be dedicated to uninformative rows of the tables. We opted for an additional row containing the keyword "Others" and the reason why they do not appear in separate rows. We also claim that adding the full list of papers in the Appendix and in the repository provides the necessary information about what *others* include. We thank the Reviewer for this useful remark; we are also open to suggestions on this.

Comment 13. Page 9, RQ5: The proposed description appears somewhat ad hoc with respect to the application of machine learning in this domain. Were automated methods that do not rely on machine learning also considered?

Response: We actually considered, in the semi-automatic approaches, also papers working with strategies that are not part of the set of ML algorithms, as shown in Figure 10(b) on page 15 in which the different strategies for the automatic part are shown. RQ_5 is still valid for this section given that existing approaches may benefit from existing ones or some datasets. We agree, however, that the description may be ambiguous and we decided to modify the description of that paragraph.

Comment 14. It is not clear how Figures 6a and 6b were generated, nor how the conclusions derived from them were obtained.

Response: We report some sentences in the paragraph RQ_8 to clarify this aspect:
"having this non-geometrical distance function, agglomerative clustering is adopted with the silhouette score [...] Figure 6(a) shows the corresponding silhouette scores, with $k = 18$ as the best. Using the best k, we can then visualise the clustered methodologies together. For this, we reduce the 8 features to 3. Principal Component Analysis (PCA) is employed [...] we exhibit the 41 methods in Figure 6(b)."
We understand that the distance between this portion of text and the figures may complicate the understanding, and also, a clear discussion on the results was missing. We introduced more information in the Discussion paragraph on page 10. In brief, the two figures are the results of the clustering proposed starting from Table 5: the first one is the silhouette score by varying k, and the latter is the visualisation (in 3D after dimensionality reduction) of the clusters. We also updated 6(b) with a clearer version.

Comment 15. Page 9, RQ8: The description provided is not sufficiently clear.

Response: Also thanks to other comments, we enriched and improved the description of RQ_8 of the second category by describing the labels and the information extracted for the classification.

Comment 16. Page 13, RQ8: The objective appears to be the classification of methodologies. Based on the results presented, would "clustering" be a more appropriate term than "classification"?

Response: We thank the Reviewer for this interesting point. Actually, clustering is involved in two of the three intra-category classifications, so it does not apply in any case to the third category. Apart from this, with "classification", we want to emphasise the goal of answering RQ_8; while, by using "clustering", we may be referring to the technical strategy adopted. As said also for other comments, the goal is to provide readers with a picture in which methodologies are related according to the necessary features for adoption. In this sense, we are interested in positioning papers rather than grouping them *tout court*, although grouping helped us in designing the pictures we aimed to propose. We appreciated this comment and are actually considering where to introduce this reflection in the paper in case of acceptance.

Semantic Web Journal
XX(X):1–28

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

From Manual to LLM-assisted: A Mapping Review on Ontology Construction Methods

Davide Di Pierro[1], Lylia Abrouk[1], Alexis Guyot[3] and Danai Symeonidou[2]

Abstract

Ontologies play a fundamental role in knowledge engineering and artificial intelligence tasks by providing a shared and formal representation of knowledge. Ontologies allow for a common understanding of concepts, support logical reasoning, and enable tasks such as inconsistency detection or instance checking. Formal ontologies pave the ground for knowledge reuse and sharing, and can be queried through dedicated query languages (SPARQL).

Constructing ontologies is pivotal to describing a domain before instantiating it or to generalising from existing data. In the literature, plenty of ontology construction methods have been proposed, according to various scenarios and data. Overall, methods can be divided according to the nature of the process: manual, semi-automatic, or automatic. This distinction is nowadays becoming blurrier with LLMs playing a primary role in tempering their distance. In this plethora, choosing a suitable methodology according to the many factors involved is demanding. The goal of this work is therefore to put into context recent updates in ontology engineering and present the necessary additional information targeted to reuse. No review of manual ontology construction has been available for the last five years, and no classification of construction methods is available. This review fills this gap by providing an updated overview and proposing a two-tier categorisation. The state-of-the-art is primarily divided into three categories (manual, semi-automatic, and automatic); for each category, representative features are then proposed for classification and comparison. To the best of our knowledge, this is the first review to offer such an intra-category classification.

Keywords

Ontology, Construction, Review, LLMs

Introduction

Ontologies lie at the intersection between computer science and philosophy. In the philosophical tradition, ontology refers to the study of what exists, the kinds and structures of objects, properties, events, processes, and relations in every area of reality Smith (2012). In this perspective, it resembles the relation between physics and metaphysics, i.e., abstracting from the physics laws to understand whether alternative models of the universe are possible. Following the ambitious objective of representing universal common sense, the Cyc ontology Lenat et al. (1985) was one of the first large efforts in ontology construction. Although still one of the richest sources for encyclopedia knowledge, its objective was too extensive, leaving the need to construct and refine ontologies for smaller domains or use cases. To fulfill the current needs of scalability, reusability, management, and control, the field of knowledge engineering Studer et al. (1998) provided a suite of guidelines and principles to design, construct, and use large-scale ontologies.

With the spread of the Semantic Web (SW) Berners-Lee et al. (1998), ontologies began to be conceived, as today, as shared conceptualisations of a (limited) universe of discussion. Thanks to the sharing of semantics through the web, many domains have shown interest in discovering and distributing knowledge across communities. Domains like chemistry, biology, and bioinformatics are today largely represented. Given their reasoning and shareability properties, ontologies are sometimes seen as an extension of databases. In order to facilitate the migration from the two technologies, much effort in the literature Hert et al. (2011); Michel et al. (2014) has been made in order to map information from one perspective to the other.

The role of ontologies is significant for both academia and industrial applications, always operating in combination with cutting-edge artificial intelligence technologies. Given their level of formality, they are capable of tracing answers or inferences, making them amenable to contexts in which explanations, transparency, and ethics are central.

With the increasing complexity of ontologies, designing efficient, sufficiently general, and easy-to-share methodologies became a primary objective for many practitioners of the field. At the beginning of the discipline, only manual ontology construction pipelines were conceived, as the state-of-the-art in machine learning was unsatisfactory for automatically constructing ontologies. However, the subsequent evolutions of machine learning did not make manual ontology research obsolete. There are still situations in which manual creation is the best (or the only) possibility. Conversely, the development of Deep Learning LeCun et al.

[1]LIRMM, Université de Montpellier, France
[2]MISTEA, INRAE & Institut Agro, France
[3]LIS, Université Aix-Marseille, France

Corresponding author:
Davide Di Pierro, LIRMM, Université de Montpellier, France.
Email: davide.di-pierro@umontpellier.fr

(2015) models, especially in the field of Natural Language Processing (NLP) Chowdhary and Chowdhary (2020), led to automatic processes in which the conceptualisation is inferred from available data (curated documents, vocabularies, etc). Thanks to the shift of paradigm, many inferential tasks became possible, reducing the demand for domain experts. Following several reviews Asim et al. (2018); Al-Aswadi et al. (2020); Zengeya and Fonou-Dombeu (2024); Armary et al. (2025), we take a glance at the performance of modern models in solving many tasks originally thought to be only human-based, like term extraction, relation extraction, axiom learning, clustering (for synonyms), and so on. As evidenced in Khadir et al. (2021), challenges in this line consist mainly of training aspects, data availability and quality, which often require a pipeline of models and synthetic data creation. Going further, Transformer models Vaswani et al. (2017) for text, e.g., LLMs Naveed et al. (2023), revived the semi-automatic literature by supporting ontology designers with broad commonsense knowledge, helping them in the execution of the manual construction phases instead of replacing human discretion. Recently, a survey on ontology construction with LLMs has been proposed Li et al. (2026), in which many works are put into correlation with the corresponding tasks of ontology management. This work highlights how LLMs can indeed fulfil several ontology-related tasks, but also shows the lack of consensus in their adoption and the limited availability of general strategies; these will be mentioned throughout. Besides, their performance is debatable in general, since benchmarking LLMs for an ontology task is still missing overall. LLMs were not only introduced in semi-automatic approaches. The availability of data (mostly in the form of documents) can be used to construct ontologies that can be proposed for the latest GraphRAG models Jiang et al. (2026). Apart from retrieval, they brought to the community advances in data fusion Huang et al. (2026), summarisation Shukla et al. (2025), validation Sesen et al. (2025), and explanation Zhang et al. (2026). Thanks to this synergy between ontologies and NLP, the two areas are developing in parallel, as the language models perform better with high-quality, validated, and consistent graphs. Finally, the database community is proposing some new research on how to better store and index property graphs Angles (2018) in order to bring benefits to GraphRAG performance.

Today, a wide variety of construction techniques address both generic and specific needs. It is generally agreed that a one-fits-all solution is not available, given the many technical parameters and constraints, not to mention the expected human values to be observed and the human skills. To support ontology practitioners in selecting a methodology, understanding similar alternatives, and being aware of the ramifications of the choice (context, metrics, weaknesses, etc), this study conducts a comprehensive review of ontology construction methods, from purely manual to LLM-supported. We bring the following contributions:

1. Through a comprehensive systematic search, we broaden and update the literature on manual (which has not been updated since 2020), semi-automatic (including LLM) and automatic ontology construction, classifying papers in their categories.
2. We answer relevant questions that affect the methodology adoption.
3. We propose an inter-category classification to help researchers in finding a methodology that fits their needs.

This review followed the Petersen et al. (2008) methodology, whose step definitions are mentioned below. The classification has been done on 136 works, and several additional reviews have been taken into account to propose classifications, metrics, and papers. The papers, resources, code, and classifications are available in a publicly accessible repository [*].

This manuscript is structured as follows: Section Background Knowledge provides the basic definitions of the involved concepts, and the necessary knowledge concerning the technologies or evaluation metrics that appear throughout the manuscript. Section Review Methodology describes and justifies our methodology, comprising the research questions, the criteria for paper selection, and some quantitative information of the selection. Section Mapping Review provides global statistical analysis, answers the research questions, and exhibits an intra-category classification among the construction methods. Finally, Section Conclusion concludes the manuscript.

Background Knowledge

In this section, we lay the ground with the basic definitions and concepts to understand the criteria to differentiate and classify the processes of ontology creation.

Ontology and Metrics

Ontology means a specification of a conceptualisation. It is a description of the concepts and relationships that can exist for an agent or a community of agents Gruber (1993). An ontology is based on a conceptualisation that is the set of objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. A conceptualisation is an abstract, simplified view of the world that we wish to represent for some purpose Genesereth and Nilsson (2012). All the definitions of ontologies put emphasis on specifying the importance of the domain of use, the specific area of interest. That is why the computer science field rejected the idea of a one-fits-all ontology. More formally, Taye (2010) defines an ontology as a 5-tuple $O := (C, R, H_C, H_R, I)$ where:

- C represents a set of concepts (or classes).
- $R \subseteq C \times C$ represents a set of (named) relations (also called properties) associated with concepts.
- $H_C \subseteq C \times C$ represents a concept hierarchy relation between concepts, indicating that the subject (domain) is a subconcept of the object (codomain).
- $H_R \subseteq R \times R$ represents a relation hierarchy relation, indicating that the first is a subrelation of the second.

[*] https://github.com/davide1797/ontology_constructions/tree/main

- I is the instantiation of the concepts, indicating, for each concept $c \in C$, which are the corresponding elements belonging to c.

R, in some contexts, is split into R_O (object properties), in which the codomain is an element of C, and R_D (datatype properties), in which the codomain is a simple datatype (string, int, date, etc). Depending on the uses, an ontology may also contain an axiomatic section for rules. Many rule formats can be described under the W3C[†] recommendations. In the SW context, rules are intended to be non-monotonic Reiter (1988) and interpreted with the Open World Assumption Moore and Van Pham (2015). Reasoning is one of the basic forms of simulated thinking, and a process of deducing new judgements (conclusions) from one or several existing judgements (premises) Chen et al. (2020). Many reasoners are available to infer new knowledge by means of explicit or implicit rules, such as Pellet Sirin et al. (2007) and Hermit Glimm et al. (2014). The most common reasoning task is consistency checking Shofi and Budiardjo (2012), the core task for correct deductive reasoning. To fully exploit reasoning capabilities, ontologies must be expressed through formal languages such as the Web Ontology Language (OWL).

Ontology metrics Lourdusamy and John (2018) are quantitative information indicating the quality of the ontology. We report here only those that will appear in the remainder of the paper:

- *Attribute Richness (AR)*: average number of attributes per class, computed as $AR = \frac{|ATT|}{|C|}$ where ATT is the set of attributes and C the set of classes.
- *Inheritance Richness (IR)*: it describes the distribution of information across different levels of the ontology's inheritance tree, computed as $IR = \frac{\sum_{C_i \in C} |H^C(C_1, C_i)}{|C|}$ where $H^C(C_1, C_i)$ is the set of subclasses of C_i.
- *Relationship Richness (RR)*: it is defined as the ratio of the number of (non-inheritance) relationships (P), divided by the total number of relationships (the sum of the number of inheritance (H) and non-inheritance (P) relationships), computed as $RR = \frac{|P|}{|H|+|P|}$.
- *Class Richness (CR)*: distribution of instances across classes, computed as $CR = \frac{|C_\epsilon|}{|C|}$ where C_ϵ is the set of classes for which no instances exist.
- *Coverage*: it answers the proportion of covered items compared to the total number. Its exact formula depends on the context. For instance, in bottom-up approaches, it may be computed as the ratio between the number of deduced facts and the number of *deducible* facts.

Apart from these quantitative metrics, a standard evaluator for ontologies is the OntOlogy Pitfall Scanner (OOPS) Poveda-Villalón et al. (2012) tool. It is a *pitfall* detector, alerting about conceptual errors in ontologies. Errors are classified as major or minor, and they are continuously updated.

Knowledge Graph

Knowledge Graph (KG) is basically a semantic network and a structured semantic knowledge base that can formally interpret concepts and their relations in the real world Xu et al. (2016). It is structurally based on simple triples in a highly flexible form. Therefore, reasoning over triplets is not necessarily limited to reasoning based on rules, but can also be diverse. We can follow the definition by Angles et al. (2017) to express a KG in terms of its graph structure. A KG can be seen as an edge-labelled graph $G := (V, E)$ where:

- V is a finite set of vertices (nodes),
- E is a finite set of edges; formally, $E \subseteq V \times L \times V$ where L is a set of labels.

A KG is generally intended as an ontology and instances, both in the form of triples. The definition of ontology can be easily translated into a KG in which H_C and H_R are sets of edges with label $\langle \texttt{isA} \rangle$, and I is a set of $\langle \texttt{instanceOf} \rangle$ edges. There are many alternatives to KG representation Hogan et al. (2021), each associated with its own language, expressiveness, and query language. While keeping their logic-based interpretation, the graph structure is suitable to infer or complete knowledge through statistical artificial intelligence Sugiyama (2015), sometimes referred to the term of *Machine Learning*.

Machine Learning, Tasks & Metrics

Machine Learning (ML) studies how to build systems that improve performance through experience Jordan and Mitchell (2015). In its most common fashion - supervised learning - training data consist of (x, y) pairs, and the goal is to predict y^* for a new x^*. The first conceived models are part of the so-called "classical machine learning", including decision trees, SVMs, neural networks, and Bayesian classifiers Hastie et al. (2009). Deep Neural Networks outperformed the classical models when dealing with larger and noisier data. A common supervised task for ontologies is link prediction Martínez et al. (2016) in which all the common ML models can be adopted Arrar et al. (2024). Performance in supervised learning is commonly evaluated using the *confusion matrix* Susmaga (2004), from which metrics such as *recall*, *precision*, *accuracy*, and *F1 score* are derived. More complex metrics are in the form of curves rather than values, like the *ROC–AUC* and *precision–recall* Erickson and Kitamura (2021) curves. *Hits@k* is suited for recommendation or ranking tasks. Variations of these metrics for ontology learning lead to the *ontological loss and improvement* metrics Sabou et al. (2005).

Unsupervised learning focuses on discovering structure without labels Murphy (2012), including clustering Rokach and Maimon (2005) and dimensionality reduction Van Der Maaten et al. (2009). Clustering is one of the most common tasks for class extraction in bottom-up approaches Yu and Hsu (2011); Mahmood et al. (2022). The quality of this task is typically assessed through cohesion and separation criteria Tan et al. (2016). Both supervised and unsupervised learning may benefit from statistical analysis (*t-test*, *p-value*, *support*, *confidence*, *Cohen's Kappa*).

Extending Deep Learning models, Transformers Vaswani et al. (2017) were conceived, based on an encoder-decoder structure Cho et al. (2014). From these architectures,

[†] https://www.w3.org/TR/owl2-primer/

Large Language Models (LLMs) Naveed et al. (2025) have emerged as cutting-edge artificial intelligence capable of generating text with coherent communication y Arcas (2022) and generalising to multiple tasks Radford et al. (2019).

The LLM's ability to solve diverse tasks with human-level performance comes at the cost of slow training and inference, extensive hardware requirements, and higher running costs. Such requirements have limited their adoption and opened opportunities for the design of better architectures Chowdhery et al. (2022) and training strategies Touvron et al. (2023). Given the need to introduce, merge, and relate semantics to terms, LLMs excel at understanding and relating general terminological knowledge. This is why they are being adopted as a support for the ontology designer in tasks like hierarchy creation, list of terms creation, and requirements elicitation. Generally speaking, LLMs are employed as a support within existing ontology construction methodologies. Assessing an LLM in general is challenging; hence, it is generally preferred to use the standard metrics in the field in which they are adopted. It is also possible to consider statistical analysis on token generation, like the *perplexity* metric. More social aspects like bias and trustworthiness are increasingly incorporated into evaluation frameworks Chang et al. (2024).

Review Methodology

This section intends to describe and justify the methodology used to conduct this review. We followed a sequential and object-oriented approach. The steps have explicit titles that help in understanding their role, but we also provide the readers with a brief description, although most of the methodology is not new. Some steps require explicit parameters that guide the process. We leverage the *Systematic Mapping Studies (SMS)* Petersen et al. (2008) approach, which involves several phases, from the definition of research questions to paper selection and classification, to answer relevant questions in the literature. The SMS approach has been adapted according to several objectives by different authors. We adapted the methodology followed in Ahaggach et al. (2024) to conduct the review. We report in Figure 1 the necessary steps:

1. **Formulating Research Questions**: The objective of the review can be summarised in these research questions, which we are going to answer. Some questions may have different interpretations according to the category we are discussing.
2. **Defining Filtering Criteria**: Keywords are the first criterion to guide the search towards the relevant papers. The basic logical operators AND and OR are generally supported by the most popular scientific search engines for complex query answering.
3. **Database Lookup**: In this phase, several sources provide the papers in compliance with the requested filtering criteria.
4. **Screening**: The first screening verifies that papers are within to the scope of this review. As an example, by searching for "ontology" and "learning", papers regarding ontologies in the educational field are returned, although they are mostly not relevant to the scope of this review.
5. **Duplicates Removal**: Duplicated papers coming from different sources are removed.
6. **Inherence Filtering**: This stage is crucial to select the papers that will be used for selection and classification. In this stage, the title, abstract, and conclusions are used to filter the works that present an innovative methodology. In other words, papers that only applied existing solutions are excluded. Moreover, methodologies that are too closely tied to specific projects and for which there is neither evidence of reuse nor citations are excluded.
7. **Selection**: In this stage, papers are sorted according to their category of construction (manual, semi-automatic, automatic).
8. **Intra-class Classification**: Every paper is classified and compared within its category. The parameters and metrics for comparison are highly dependent on the corresponding category.

Formulating Research Questions

To conduct this study, it is imperative to specify which questions we aim to answer. Questions should provide enough information to categorise ontology construction methods, clarify the reasons for which they have been conceived, and clarify their advantages and disadvantages. A goal of this contribution is to provide sufficient information for practitioners when selecting a construction methodology. To facilitate the reuse of a methodology, it is fundamental to known how it has been evaluated (RQ_4), if it has origin from another methodology or is completely novel (RQ_5), the contexts in which it is applicable (RQ_6), and its limitations (RQ_7). Apart from this, we can come across several methodologies presenting similar scenarios of use. This is why we also propose an internal classification (RQ_8) within the three categories of ontology creation approaches. To do this, a preliminary split into the three categories (RQ_3) is necessary. We also propose global statistics of this research trend (RQ_1 and RQ_2).

Hence, the proposed questions are the following ones:

- RQ_1: What is the annual number of methodologies proposed?
- RQ_2: In which conferences/journals have they been proposed?
- RQ_3: How to (globally) classify the methodologies?

For every category:

- RQ_4: What metrics are available?
- RQ_5: Is there any starting methodology/dataset adopted?
- RQ_6: What is the context of the methodologies?
- RQ_7: What are the limitations of the methodologies?
- RQ_8: How to classify the methodologies?

Defining Criteria Filtering

The objective in this step is to define which are the keywords that are linked to the set of relevant papers. The main sources of information are the most renowned academic databases like Scopus, Scholar, and IEEE Xplore. We decided to analyse two sets of papers: (i) from the last 10 years for

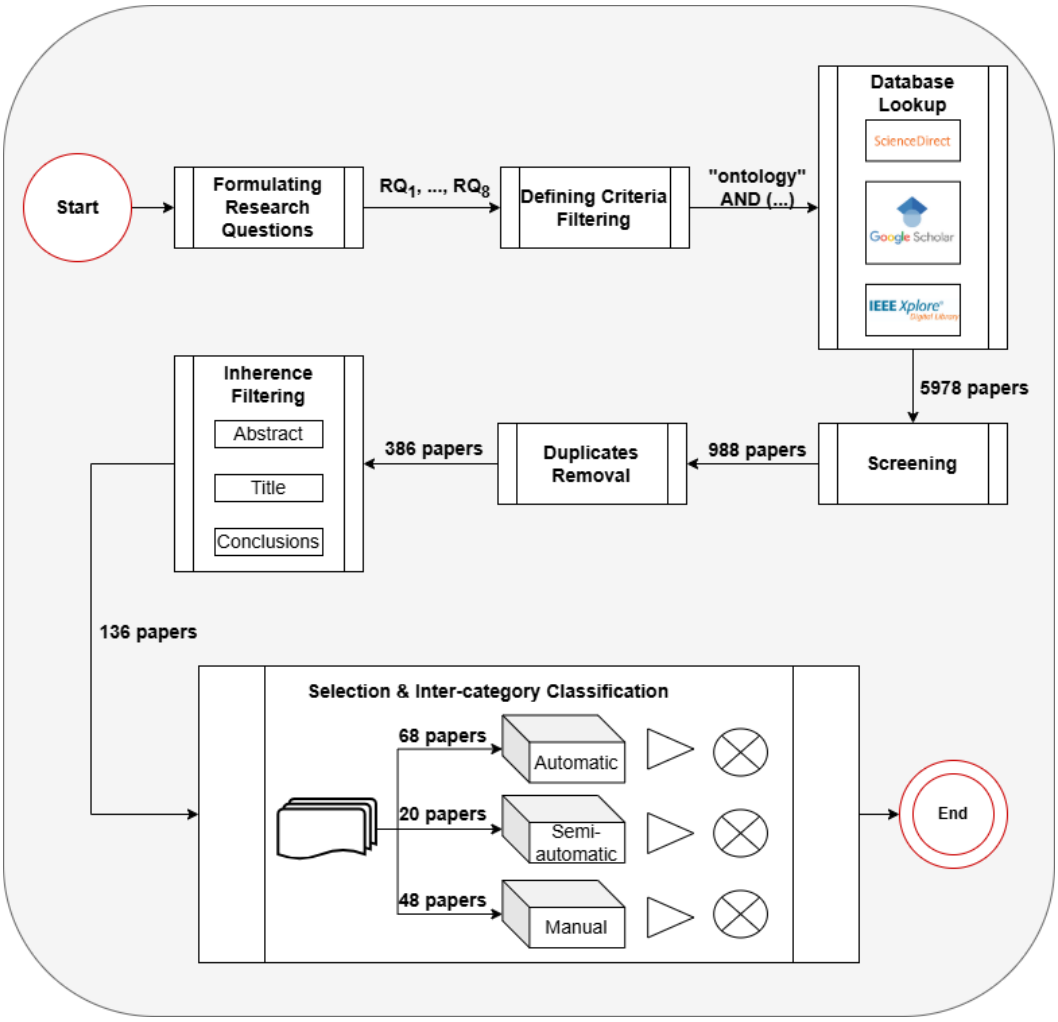


Figure 1. Chosen review methodology with the different steps and number of input/output papers for every step. In the last phase, papers are split according to the category and classified accordingly.

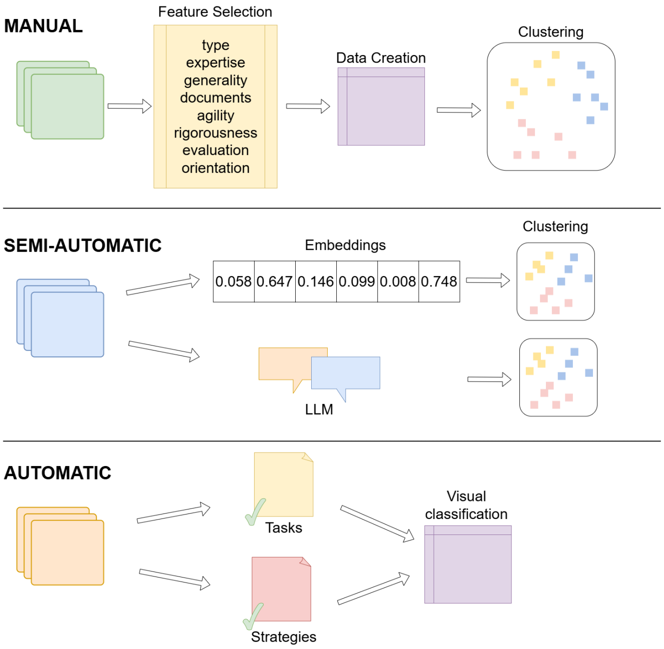


Figure 2. Intra-category classification methodologies.

ontology learning, and (ii) manual ontology construction with no time filter. We used a string pattern to craft relevant papers based on logical operators. The chosen queries, applied to titles, are shown in Listing 1:

Listing 1: Search query patterns for (i) and (ii)

```
(i)  "ontology" AND ("methodologies" OR "
    construction" OR
     "methodology" OR "method" OR "creation"
        OR "learning" OR "LLM")

(ii)  "ontology" AND ("methodologies" OR "
    methodology" OR "method")
```

Database Lookup

In this phase, the defined query must be used to retrieve relevant papers in the most popular academic search engines. We searched the papers in Science Direct[‡], Google Scholar[§], and IEEE Explore[¶]. All of them provide intuitive ways to use the logical operators AND and OR, by using the same formalism, very close to the shown query pattern.

‡ https://www.sciencedirect.com/
§ https://scholar.google.com/
¶ https://ieeexplore.ieee.org/Xplorehelp/searching-ieee-xplore/search-engine

Reviews are particularly taken into account in order to fetch papers that might be missed by our definition of criteria. For these search engines, an API is available to retrieve the basic information (title, authors). We also need additional information, e.g., the abstract (for further filtering) or the number of citations. Unfortunately, this information could not have been retrieved automatically for every paper; hence, an exhaustive manual check was required. In the end, *5978 papers* were returned.

Screening

During screening, the first (and major) selection is performed. We asked an LLM (GPT 3.5) to look at all the titles and to highlight which papers were, with high confidence, out of the scope of this review. We explicitly asked to keep the surveys. Provided with a few prompts of examples, the model was able to separate many non-relevant papers. Then, we removed some papers or proposals published in poorly recognised venues. In this phase, we were much more interested in avoiding false negatives rather than false positives. This is why, for the remaining papers, we opted for manual filtering, looking at the titles and abstracts. In the end, about 1/6 of the initial dataset was kept, leading to *998 remaining papers*.

Duplicates Removal

Many papers are expected to overlap across the different sources. For this reason, duplicates have been removed. We also removed similar papers in which one is the extension of another, without introducing novelty in ontology construction, finally leading to *386 distinct papers*.

Inherence Filtering

Finally, a more semantic filtering is necessary in order to capture those works that introduce novelty in the scope of the review, i.e., proposing a new methodology for ontology construction or learning. Filtering is performed by examining first the abstract, title, and conclusions, and, if necessary, the entire content. In order for a paper to be considered *in line* with the scope of the survey, the following criteria were taken into account: *goal*, *novelty*, and *reproducibility*. For *goal*, we mean that the work actually solves a problem of ontology construction/learning, and so excluding related (but distinct) lines like population, enrichment, and so on. There are, evidently, papers working for more tasks, but construction or learning is mandatory in order to be included; for *novelty*, we mean that the work is clearly positioned in the state-of-the-art and does not represent a smaller improvement of an existing work; finally, for *reproducibility*, we mean that the workflow is well-described in its steps, and the adopted datasets are standard or well-described. Review papers in this phase are excluded, although they have been used to retrieve or consider possible relevant papers not returned by the query. Non-accessible papers were excluded in this phase as well. In the end, *136 papers* have been selected for the final comparison. This part was inherently manual and time-consuming, but it took advantage of the first filters that shrank the workload.

Selection & Evaluation

Each paper has to be sorted according to the category it belongs to. Distinguishing among manual, semi-automatic, and automatic constructions facilitates an intra-category evaluation based on the methodologies within the same group. The *136 papers* have been split with these frequencies: *68 manual*, *20 semi-automatic*, and *48 automatic*. Intra-category evaluations are based, when possible, on features from the literature (especially reviews) and/or others provided in this manuscript. Withing each category, this contribution aims to classify papers according to its scenarios of use, requirements, limitations, etc. We refer to these conditions that impact the adoption of every methodology as generic "features". The selected metrics for every category are described later. Given the different nature of the three categories, the classification of the selected papers has been designed differently. Indeed, for the "manual" category, features concern more engineering processes, steps, or human interactions; the "semi-automatic" one refers to the chosen algorithm/strategy followed for the automatic/manual part and, finally, the "automatic" one concerns more the possible tasks and the types of suitable models to deal with them. To the best of our knowledge, no existing review on this topic proposes a quantitative/visual classification of the methodology to highlight similar works based on features concerning reusability.

In Figure 2, we provide a more detailed description of how the intra-category classification is provided for the three categories:

- **manual**: by taking inspiration from existing surveys Fernández-López and Gómez-Pérez (2002); Cristani and Cuel (2005); Jones et al. (2007); Dahlem and Hahn (2009); Iqbal et al. (2013); Nefzi et al. (2014); Sattar et al. (2020), we reused some existing features like *agility* or *orientation*, introduced new ones like *type* or *documents* and filled, for every selected paper, the values for all the 8 features. Starting from this created dataset, we propose a clustering based on Gowen's distance and the silhouette score to find the best k. Then, to provide a visualisation of the clustering, we propose a dimensionality reduction to synthesise the 8 features into 3 and visualise the clusters as a 3D plot.
- **semi-automatic**: we distinguish, for every paper, the technique adopted for the manual and automatic parts. Then, we propose two classification strategies: (i) based on embedding similarities starting from textual features leveraging a sentence transformer, and (ii) based on Huang and He (2024), in which an LLM behaves like a classifier. A discussion of the similarities and differences of the results is also presented.
- **automatic**: we distinguish, for every paper, the task(s) that it aimed to solve, and the adopted strategies. The visualisation as a 2D chart makes it evident for a reader to understand what the relevant papers dealing with a specific task are or, conversely, which are the tasks for which a strategy has already been adopted successfully.

Mapping Review

Before presenting the three core sections, we provide some basic statistical analysis by looking at how the production of approaches evolved through time. Hence, we report in this section the following metadata and charts to provide preliminary insights into the state-of-the-art for the scope. We report the following information:

- **Global number of articles per year**: number of selected articles for this review for every year (Figure 3).
- For every category, **number of articles per year**: illustrates the publication trend of publications according to the three categories from 2016 (Figure 4).
- for every category, **global number of citations per year**: computed as the sum of all paper citations within the same category through the years from 2016 (Figure 5)
- **Most popular paper venues**: Figure 6.

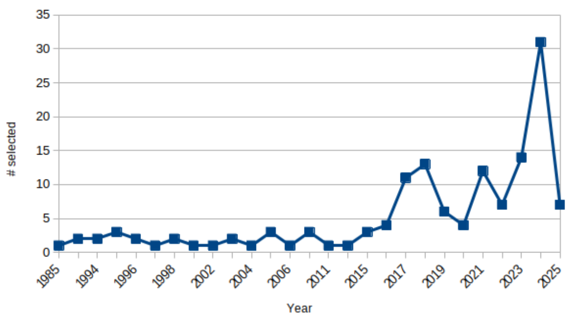


Figure 3. Number of selected publications per year.

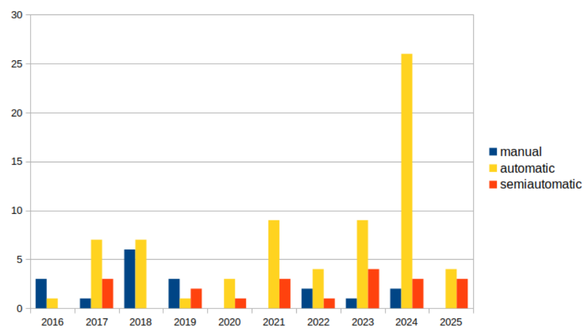


Figure 4. Number of selected publications per year and category from 2016.

We split this section according to the categories in which we divide the state-of-the-art in ontology creation (i) manual, (ii) semi-automatic, and (iii) automatic. The (i) one is often associated with the historical methodologies, but this research line still motivates new fundamental work since there are many scenarios in which relying on manual creation is the best (or unique) option. The (ii) one is the least represented, but it is gaining momentum thanks to the introduction of LLMs in many steps of the construction processes. Finally, the (iii) one progresses through the main ML models and advancements. For the categories (ii-iii),

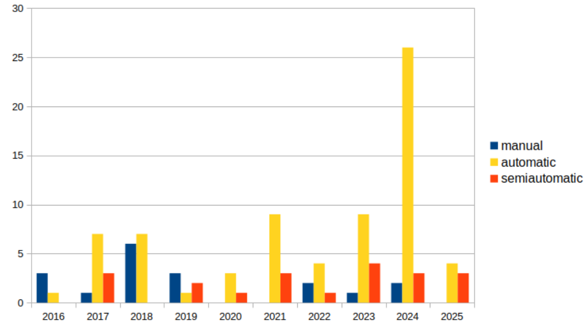


Figure 5. Number of citations per year and category from 2016.

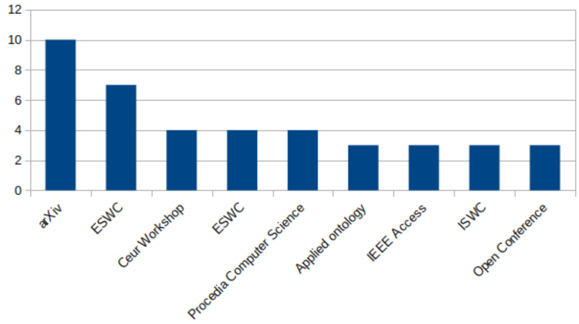


Figure 6. Most popular paper venues.

only the last ten years (2016-2025) are included in this review, since there is a clear benchmarking in ML indicating that new architectures overcome the previous ones. For the three categories, the given classification is based mainly on the reuse, i.e., under which conditions the methodology can be applied.

Manual Ontology Creation

Manual ontology creation is a research area at the intersection of artificial intelligence and knowledge engineering. Over the decades, many methodologies have been proposed to design sound and complete conceptualisations, capable of representing the complexity of a domain without introducing fallacies or inconsistencies. Many techniques borrowed the methodological steps from the software engineering field, which has evolved from the traditional waterfall development Adenowo and Adenowo (2013) to the modern agile-based Thesing et al. (2021) one. Over the years, methodologies have originated for several reasons. In many cases, there was a specific interest in a domain for a project or industrial application. In these situations, mostly, a specific methodology has been developed to solve the problem, but there are examples of methodologies that were followed in other contexts, too. With the introduction of Linked Data and the availability of much more shared and interoperable information on the Web, some ontology construction methods have been tailored to comply with these technologies and prioritise the Semantic Web representation. This contribution provides a clustering of the methodologies based on the proposed distinguishing features. To the best of our knowledge, the last survey on manual ontology creation was done in Sattar et al. (2020), and the state-of-the-art can be updated in the field. Apart from updating the literature, the two main gaps with respect to existing reviews are:

- providing a classification specifying the necessary context of usage (e.g., if it is general or specific to a sector, if it requires people educated in formal languages or not, etc.),
- providing a clustering of methodologies based on the context of usage.

RQ_4: What metrics are available? The comparison between manual methodologies cannot be performed on some quantitative results of the creation process, given that all of them may converge to the same result. Henceforth, we look for metrics that refer to the creation process itself, and by its capability of being reproducible, explained, or generalised. Some features are taken from existing surveys; others are proposed by us. We report in Table 1 the detected features (when available) of the selected reviews on manual ontology construction. We highlight in the Table that the metrics are almost all related to the software engineering process and that, in recent years, as evidenced by the last cited paper, more features are related to the LOD and FAIR principles, so aspects like reusability, localisation, and interoperability become imperative.

RQ_5: Is there any starting methodology/dataset adopted? Different methodologies take inspiration from existing, stable, or popular ones. The majority of the works extend earlier methodologies, adapting them to modern needs. These extensions cannot easily be inferred solely by looking at the titles or names, with the exception of UponLite De Nicola and Missikoff (2016) being the extension of Upon De Nicola et al. (2005) with the same authors. We report in Table 2 the list of works with their corresponding extended methodologies. We took into account only methods in the ontology development field, i.e., methods taking inspiration from the software engineering field are beyond our scope. Here, it is evident how Methontology represented a milestone in the field, although recent works also extend agile methodologies like NeOn or LOD.

RQ_6: What is the context of the methodology? Each methodology may have a specific context or domain of application. Some of them aim to be general enough to be reused without constraints, while others depend on external conditions. We report in Table 3 the methodologies explicitly specifying their corresponding contexts of use. More generic constraints, such as those generally agreed upon within top-down, bottom-up, or linguistic approaches, are not specified here, as RQ_8 will clarify these aspects.

RQ_7: What are the limitations of the methodology? Most common methodologies have been extensively studied, and frequently constitute a basis for subsequent methodologies. For this reason, their limitations are also more evident, given the need, for instance, to add validation phases, iterative steps, etc. Table 4 lists the most relevant reported limitations. We specify that we mention only explicit limitations provided by the authors in the papers or by authors of papers who extended the cited one. Here again, it is evident that only recent methodologies are suitable for current needs like reproducibility and integration. From this point of view, the FAIR principles have revolutionised the field, and earlier methodologies can be used more as a starting point rather than a complete solution.

RQ_8: How to classify the methodologies? The proposed methodologies have been selected from the literature according to some general criteria like their capability of being applied in multiple scenarios, their impact on the state-of-the-art (even if not strictly generic), their historical value, or their influence on the next strategies. We selected surveys on the topics, ontological construction methodology papers, and ontological conceptualisation papers. Among the possible features, we propose ten that provide an immediate overview of the context in which the methodology can be employed. The features do not have a preferred value; they are only intended to be descriptive. It is the specific scenario that helps us identify whether a value for a specific feature is more suitable than others. The chosen features are:

- **type (F1)**: it specifies the origin and the reason for which a methodology has been proposed, providing also hints about how to reuse it. Types are distinguished into *project* (p), *framework* (f), *integration* (i), and *linguistic* (l). p refers to methodologies born in the context of the development of a project, often presenting technical specificities but sometimes generalisable; f refers to general guidelines or proposals aiming to guide towards correct engineering solutions; i refers to strategies particularly tailored to integrating and reusing existing ontologies rather than re-engineering; l refers to the linguistic domain that comprises its own requirements and complexities, fully exploiting NLP solutions. We claim ontology construction in linguistics is neither limited to an application nor aiming to be a general framework. However, it is not only an integration, given that it also needs to understand which tasks are to be executed, the design of semantics alignment, how to guarantee the presence of the target language(s) and other possible factors. This is why we claim that linguistics deserves its own value in that feature.
- **expertise (F2)**: it specifies whether the methodology can be used by novice users (n) or requires a certain experience in knowledge representation (y).
- **generality (F3)**: it indicates whether the methodology is easily *reusable* in other domains (y) or too *tight* to one context to move into others (n).
- **documents (F4)**: it indicates whether the presence of existing documentation or resources is necessary (y) in order to develop the process or not (n). This feature is sometimes obvious, as for bottom-up approaches.
- **agility (F5)**: it specifies whether the methodology tends towards a waterfall approach (n) or an iterative (agile) one (y). Although agile software development is way more recent, in the field of knowledge engineering, many current methodologies are still waterfall-based.
- **rigorousness (F6)**: it indicates whether there is a strong degree of *formality* (e.g., OWL) during the process (y) or the formality is *hidden* by graphical tools or expressed in natural language (n).
- **evaluation (F7)**: it indicates whether the methodology includes any kind of *validation* phase by the involved stakeholders (y), or if it happens only after the development (n).

Review	Classification Features
Fernández-López and Gómez-Pérez (2002)	Project Management processes, Pre-development processes, Requirements process, Design process, Implementation process, Post-development process, Integral processes.
Cristani and Cuel (2005)	Input, Description, Output, Orientation.
Jones et al. (2007); Nefzi et al. (2014) Saad and Shaharin (2016)	Methodology phases.
Dahlem and Hahn (2009)	Adequate terminology, Structure, Descriptiveness, Transparency, Error avoidance, Robustness, Lookahead, Consistency, Hiding formality, Expressiveness, conceptualisation flexibility, Ontology assumptions, Tool support.
Badr et al. (2013)	Methodology phases.
Iqbal et al. (2013)	Type of development, Collaborative construction, Reusability support, Degree of application dependency, Life cycle recommendation, Strategies for identifying concepts, Methodology details, Interoperability support.
Sattar et al. (2020)	Domain Analysis, conceptualisation, Level of detail, Collaborative construction, Implementation, Evaluation, Instatiation, Maintenance, Documentation, Ontology localization, Support for reusability, Support for integration and merging, Support for interoperability, Estimation of human resources, Sample Application, Methodology rooted in established methodologies.

Table 1. Summary of manual ontology creation review and features.

Paper	Paper(s) Extended
Bautista-Zambrana (2015)	Methontology Fernández-López et al. (1997)
Saad and Shaharin (2016)	Methontology, Sandkuhl et al. (2005)
UponLite De Nicola and Missikoff (2016)	Upon De Nicola et al. (2005)
OmMas Yunianta et al. (2017)	Moma Ying (2009)
OntoDi Yunianta et al. (2017)	Badr et al. (2013)
Ahmad et al. (2018)	Badr et al. (2013)
Scim John et al. (2018)	Methontology
Jacksi (2019)	Methontology
Linked Open Terms Poveda-Villalón et al. (2022)	Methontology, NeOn Suárez-Figueroa et al. (2011)
Tang et al. (2023a)	Uschold & King Uschold and King (1995), Methontology
Tailhardat et al. (2024)	Linked Open Terms Poveda-Villalón et al. (2022)
Others	No papers extended.

Table 2. Summary of manual methodologies extensions.

Methodology	Context
PhysSys Borst et al. (1995)	Physics domain
Saad and Shaharin (2016)	Lesson plan domain
Moma Ying (2009) & OmMas Yunianta et al. (2017)	Multi-agent domain
Daponte and Falquet (2018)	Availability of the Dolce Gangemi et al. (2003) and OMDoc Kohlhase (2000) ontologies.
Linked Open Terms Poveda-Villalón et al. (2022)	Linked Data resources
Olivares-Alarcos et al. (2022)	Robotics domain
Sharma and Jain (2024)	Medical domain
Others	Unspecified context.

Table 3. Summary of papers' contexts for manual construction.

- **orientation (F8)**: it indicates whether the methodology is *top-down* ($\bigtriangledown$), *bottom-up* ($\bigtriangleup$), or middle-out ($\odot$).

These features are collected for the various selected papers. To do this, we checked explicit, implicit, and reference information to fill Table 5. After filling in the values for every methodology, it is possible to cluster the methodologies based on the selected features. Methodologies within the same cluster can be used under similar circumstances. A distance measure is straightforward to compute by looking at the features. F2-F7 can be considered boolean values, while F8 values are assigned this way: 0 for $\bigtriangledown$, 1 for $\odot$, and 2 for $\bigtriangleup$. In this way, the *middle-out* approach is correctly positioned in between the two methodologies, and the distance between the two opposites is maximised. Finally, the rest of the features are categorical and, hence, only equality (0) or inequality (1) is evaluated. For this purpose, we applied the Gower's distance Gower (1966)

Methodology	Limitation(s)
IDEF5 Peraketh et al. (1994)	More focused on the evaluation rather than construction Dutta et al. (2015).
Grüninger & Fox Gruninger (1995)	More focused on the evaluation rather than construction Dutta et al. (2015), not suitable for common data integration Yunianta et al. (2017).
Uschold & King Uschold and King (1995)	Not suitable for common data integration Yunianta et al. (2017).
Swartout et al. (1996)	Not suitable for common data integration Yunianta et al. (2017).
Methontology Fernández-López et al. (1997)	Interactive and visual tools are not supported De Nicola et al. (2005), tt may lead to misunderstandings and inconsistencies if the designer is not involved into the domain Gahleitner et al. (2005), more focused on the evaluation rather than construction Dutta et al. (2015), not suitable for common data integration Yunianta et al. (2017), gaps with Linked Data needs Poveda-Villalón et al. (2022).
Ontology Development 101 Noy et al. (2001)	It does not offer instructions on conceptualisation Bautista-Zambrana (2015), gaps with Linked Data needs Poveda-Villalón et al. (2022).
On-To-Knowledge Sure et al. (2003)	It may lead to misunderstandings and inconsistencies if the designer is not involved into the domain Gahleitner et al. (2005), gaps with Linked Data needs Poveda-Villalón et al. (2022).
Diligent Pinto et al. (2004)	Not suitable for common data integration Yunianta et al. (2017), gaps with Linked Data needs Poveda-Villalón et al. (2022).
Upon De Nicola et al. (2005)	It fails to provide comprehensive details for collaborative ontology construction aspect John et al. (2018).
Moma Ying (2009)	It only partially successful in enabling the development of a complete application development by the domain experts, it still requires the agent developer and the software engineer to get involve, it is tested and evaluated specifically in the financial services, the ontology can become inconsistent during the editing process, no iteraration occurs in the development Yunianta et al. (2017).
NeOn Suárez-Figueroa et al. (2011)	No ontology evaluation and validation to check the consistency aspect of the ontology knowledge, not suitable for common data integration Yunianta et al. (2019).
Badr et al. (2013)	Not suitable for common data integration Yunianta et al. (2017).
Samod Peroni (2016)	Precise recommendations for conceptualisation, documentation, tools to be used, etc. are out of the scope Poveda-Villalón et al. (2022).
OmMas Yunianta et al. (2017)	It has too many phases (i.e., nine phases altogether), which can make it less efficient Yunianta et al. (2019).
Others	No explicit limitations.

Table 4. Summary of papers' limitations.

defined in this way:

$$S_{i,j} = \frac{\sum_{k=1}^{p} w_{i,j,k} s_{i,j,k}}{\sum_{k=1}^{p} w_{i,j,k}}$$

with $s_{i,j,k} = 1 - \frac{|x_i - x_j|}{R_k}$ with R_k being the range of the k-th variable and thus ensuring $0 \leq s_{i,j,k} \leq 1$, and $w_{i,j,k} = 1$ in our case for every feature.

Having this non-geometrical distance function, agglomerative clustering Ackermann et al. (2014) is adopted with the silhouette score Shahapure and Nicholas (2020) used to find the most prominent number of clusters (k). Candidates k are from 2 to half of the number of methodologies + 1. Figure 7a shows the corresponding silhouette scores, with $k = 18$ as the best. Using the best k, we can then visualise the clustered methodologies together. For this, we reduce the 8 features to 3. Principal Component Analysis (PCA) Wold et al. (1987) is employed for dimensionality reduction. After the reduction, we exhibit the 41 methods in Figure 8, divided by clusters identified by colours. For the textual list of the cluster division, see Appendix.

Discussion The proposed clustering aims to combine methodologies that can be adopted within similar contexts or under the same conditions. Selecting among a number of methodologies can be time-consuming, and the final choice is likely to be one of the most popular approaches, overlooking possible alternatives. We claim that this classification is beneficial for those searching for an approach for their requirements, with all the main distinctive characteristics represented. We claim that, conversely to automatic approaches, these are less affected by technological evolution, and that is why, for instance, *Methontology* is still among the most used. This justifies our solution of not filtering by age in the presented approaches. This clustering is highly sensitive to the choice of k. Although $k = 18$ is the one providing the highest silhouette score, the presence of singletons, for instance, is not sufficiently informative for those seeking to identify methodologies similar to the singleton value. The scope of the visualisation for the best k is to highlight that the best split generally fulfils the task of separating methodologies.

Methodology	F1	F2	F3	F4	F5	F6	F7	F8
Cyc Lenat et al. (1985)	p	y	y	y	n	y (partly)	y (partly)	▽
Kads Wielinga et al. (1992)	f	y	y	n	y	n	n	▽
Tove Fox (1992)	f	y	y	n	n	y	n (poorly)	▽
Idef5 Peraketh et al. (1994)	f	n	y	y	n	n	y	▽
Plinius van der Vet et al. (1994)	p	y	n	y	n	y	n	△
Grüninger & Fox Gruninger (1995)	i	y	n	n	n	y	y	▽
PhysSys Borst et al. (1995)	p	y	n	y	n	y	n	▽
Uschold & King Uschold and King (1995)	f	y	y	n	n	y	y	▽
Onions Gangemi et al. (1996)	i	y	y	y	n	y	n	⊙
Sensus Swartout et al. (1996)	i	y	n	y	n	y	n	△
Methontology Fernández-López et al. (1997)	f	n	y	n	n	y (partly)	y	▽
(KA)^2 Benjamins and Fensel (1998)	i	y	y	y	n	n	n	⊙
Mikrokosmos O'Hara et al. (1998)	l	y	n	y	n	y	n	△
Ontology Development 101 Noy et al. (2001)	f	n	y	n	n	y	n	▽
eXtreme Hristozova and Sterling (2002)	f	y	y	y	y	n	y	⊙
Dolce Gangemi et al. (2003)	l	y	y	y	n	y	y	▽
On-To-Knowledge Sure et al. (2003)	p	y (partly)	y	y	y	y	y	⊙
Diligent Pinto et al. (2004)	i	y (partly)	n	n	n	y (partly)	y	⊙
DynamOnt Gahleitner et al. (2005)	i	n	y	n	n	y	n	▽
Upon De Nicola et al. (2005)	f	y	y	n	n	y (partly)	n	⊙
Sandkuhl et al. (2005)	i	y	y	y	n	n	n	△
Hcome Kotis and Vouros (2006)	i	y	y	y	n	n	n	⊙
Dogma Jarrar and Meersman (2009)	i	y	y	y	n	n	n	▽
Li et al. (2009)	f	y	y	y	n	y	y	△
Moma Ying (2009)	p	y	n	n	n	n	n	▽
NeOn Suárez-Figueroa et al. (2011)	i	y (partly)	y	y	y	n	n	△
Debruyne and Meersman (2012)	l	y	n	y	n	n	n	⊙
Grounded Ontology Nabi and Asif (2015)	i	y	y	y	n	n	n	△
Bautista-Zambrana (2015)	l	y	n	y	n	n	n	⊙
Yamo Dutta et al. (2015)	f	y	y	y (partly)	n	y	y	△
UponLite De Nicola and Missikoff (2016)	f	n	y	n	n	y (partly)	n	⊙
Samod Peroni (2016)	f	n	y	n	y	y	y	⊙
Saad and Shaharin (2016)	f	y	y	y	y	y	y	⊙
OmMas Yunianta et al. (2017)	p	y	y	n	y	y	y	▽
Ahmad et al. (2018)	p	y	y	n	n	y	y	▽
Daponte and Falquet (2018)	i	y	n	y	n	y	y	△
Fawei et al. (2018)	p	y	n	y	n	y	y	△
Scim John et al. (2018)	f	y	y	n	y	y	n	▽
Li and Alian (2018)	f	y	y	y	n	y	y	▽
Zhang et al. (2018)	p	y	y	y	y	y	y	▽
Abdelghany et al. (2019)	f	y	y	n	y	y	y	⊙
Jacksi (2019)	p	y	n	n	n	y	n	▽
OntoDi Yunianta et al. (2019)	i	y	y	y	n	y	y	▽
Olivares-Alarcos et al. (2022)	p	y	y	y	y	y	y	▽
Linked Open Terms Poveda-Villalón et al. (2022)	i	n	y	n	y	y	y	⊙
Tang et al. (2023a)	p	y	n	y	n	y	y	▽
Oda Sharma and Jain (2024)	p	y	y	y	n	y	y	▽
Tailhardat et al. (2024)	i	y	y	y	n	y	y	▽

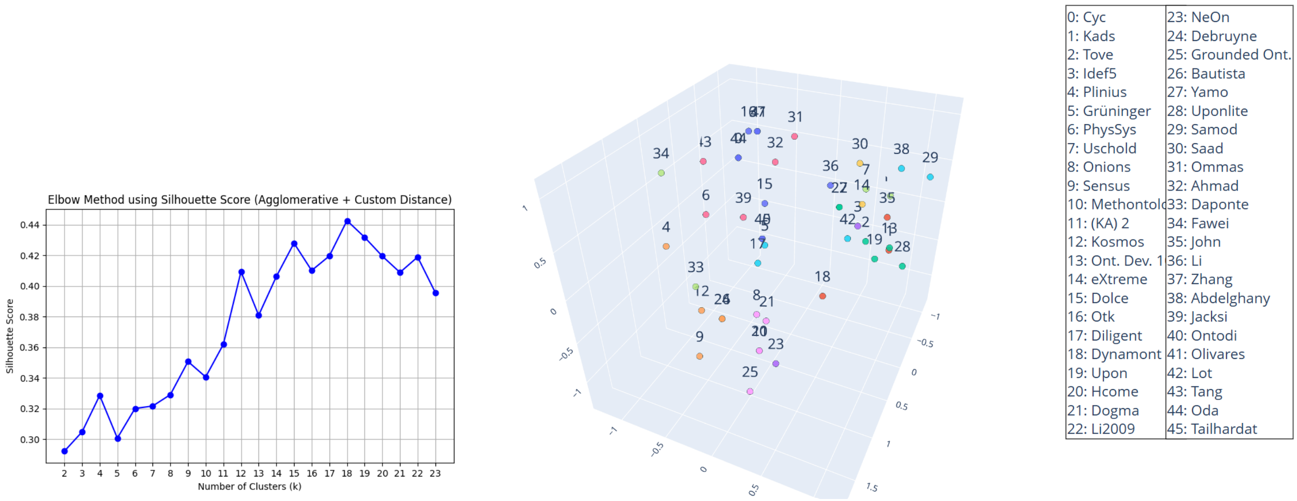
Table 5. Summary of manual ontology creation methodologies with their corresponding features F1-F8, used as the dataset for the clustering classification.

For alternative splits, we refer the reader to the repository[‖], in which the splits from $k \in [11, 20]$ are shown. The most evident limitation of this classification is the dependence on the selected features, although we claim they are representative of the current and historical needs in the field, as also supported by the literature.

Semi-automatic Ontology Creation

Semi-automatic ontology construction consists of introducing learning steps or, with the recent developments, LLMs to assist manual ontology steps. The evolution of this research

‖ https://github.com/davide1797/ontology_constructions/tree/main/data



(a) Silhouette Scores for the Different *K*s.

(b) 3D Clustering Chart of Manual Methodologies.

Figure 7. Comparison of Silhouette Scores and 3D Clustering.

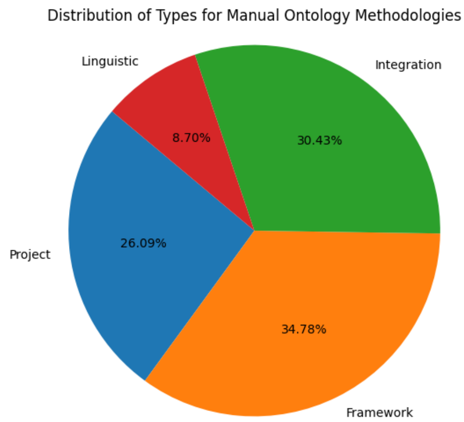


Figure 8. Distribution Pie Chart of Manual Methodologies for F1.

line has benefited from both the evolution of ML and ontology engineering. From an architectural point of view, many methodologies differ more in the automatic steps involved in the pipeline. Apart from answering the RQs listed above, we also provide a visualisation of the most common manual and automatic approaches adopted in this line in recent years.

RQ$_4$: What metrics are available? This family indeed combines metrics from traditional ontology construction and ML ones. Most common metrics are based on the traditional performance table for classification tasks or the ones in ontology engineering, such as OOPS and coverage, when a ground truth exists. We report in Table 6 the list of metrics for the semi-automatic papers. As evidenced in the Table, the metrics are almost all aligned with the traditional ML ones, overlooking the engineering process that is, however, present. The OOPS evaluation is more concerned with the engineering aspect, evaluating the potential reusability and integration of the resource. Many recent works introducing LLMs still have problems in identifying generic and reusable metrics for the model part. The lack of reproducible metrics is also one of the reasons why, to date, a shared methodology for LLMs is not yet available.

RQ$_5$: Is there any starting methodology/dataset adopted? Semi-automatic methodologies rely on pattern extraction techniques for the automatic part, which require existing data sources to infer or derive knowledge. We report in Table 7 the resources used for the approaches. In most cases, the dataset used is not known in the literature. Generally, these datasets are not large enough to train a fully automatic pipeline or the context (as for the Medical Guideline Ontology) is too specific.

RQ$_6$: What is the context of the methodology? As for manual ontology, semi-automatic construction may be specifically tailored to a domain. The specificity of this research line appears to be common, given the number of *non-standard datasets* adopted or the evaluation based on experts' revision. We report in Table 8 the context (when available) of the proposed approach.

RQ$_7$: What are the limitations of the methodology? In this section, we report the explicit (mentioned by the authors themselves) limitations of the proposed works, mostly inherent to the adopted architecture. They are shown in Table 9. Also present in this category is the lack of correct evaluation when LLMs are involved. Above that, the difficulty in generalising across different domains and datasets is also evident, as it is part of the motivation for choosing this class of approaches.

RQ$_8$: How to classify the methodologies? We propose here two classifications of semi-automatic methodologies based on the *semantic* similarity between the manual and automatic parts. For this classification, for every selected paper, we identified the manual and automatic parts adopted in the methodology. Then, we identified several common labels to clearly distinguish them. The harmonisation has been done manually based on the definition provided below. The statistical presence of these labels is shown, respectively, in Figures 10a and 10b. The definitions for every label are the following ones:

Metric	Paper(s)
Precision	Qiu et al. (2017); Fawei et al. (2019); Doumanas et al. (2024); Wu et al. (2025)
Recall	Qiu et al. (2017); Fawei et al. (2019); Doumanas et al. (2024); Wu et al. (2025)
Accuracy	Chandolikar et al. (2019); Fawei et al. (2019); Dera et al. (2020); ten Haaf et al. (2021); Sobhgol et al. (2023)
Welch t-test	ten Haaf et al. (2021)
Support, confidence	Zhao et al. (2021)
OOPS, Coverage	Fawei et al. (2019); Mesmia and Mouhoub (2023); Pan et al. (2023)
F-Score	Fawei et al. (2019); Doumanas et al. (2024)
Non-standard metrics	Rani et al. (2017)

Table 6. Metrics for semi-automatic ontology construction approaches.

Dataset	Paper(s)
SemEval-2016	Dera et al. (2020)
Medical Guideline Ontology	ElAssy et al. (2022)
Enterprise Model Fox (1992)	Pan et al. (2023)
Arabic WikiPedia	Mesmia and Mouhoub (2023)
New dataset	Qiu et al. (2017); Rani et al. (2017); ten Haaf et al. (2021); Zhao et al. (2021); ElAssy et al. (2022); Mesmia and Mouhoub (2023); Pan et al. (2023); Tang et al. (2023b); Huang et al. (2024b)
Non-standard datasets	Qiu et al. (2017); Wang et al. (2017a); Chandolikar et al. (2019); Fawei et al. (2019); Dera et al. (2020); Ma et al. (2021); Sobhgol et al. (2023); Huang et al. (2024b); Kommineni et al. (2024); Gadusu et al. (2025); Lichtnow et al. (2025); Wu et al. (2025)

Table 7. Datasets for semi-automatic ontology construction approaches.

Context	Paper
Chinese Tax	Qiu et al. (2017)
Topic Modelling	Rani et al. (2017)
E-books	Wang et al. (2017a)
Marathi Language	Chandolikar et al. (2019)
Law	Fawei et al. (2019)
Sentiment Analysis	Dera et al. (2020)
	ten Haaf et al. (2021)
Relational databases	Ma et al. (2021)
Restaurant	ten Haaf et al. (2021)
Power	Zhao et al. (2021)
Medicine	ElAssy et al. (2022)
Linguistic	Mesmia and Mouhoub (2023)
Energy	Pan et al. (2023)
Manufactury	Sobhgol et al. (2023)
Autonomous driving	Tang et al. (2023b)
Water distribution	Huang et al. (2024b)
Health	Gadusu et al. (2025)
Smart Campus	Lichtnow et al. (2025)
Requirements elicitation	Wu et al. (2025)
Unspecified context	*Others*

Table 8. Contexts of semi-automatic ontology creation papers.

- **manual**: *Creation and experts' revision*: it involves the definition of concepts by exchanging opinions with experts; *Middle-out*: it follows a middle-out approach, starting from some basic concepts and then spreading; *Scraping*: it relies on existing sources on the web that can be analysed automatically; *Relational database mapping*: it maps data directly from a relational database; *Ontology Development 101*: it follows the Ontology Development 101 methodology; *NeOn*: it follows the NeOn methodology; *HCOME*: it follows the HCOME methodology; *UponLite*: it follows the UponLite methodology; *McGinty et al.*: the methodology follows the McGinty (2018) approach.
- **automatic**: *Spanning Tree*: it relies on spanning tree algorithms for hierarchy creation; *Singular Value Decomposition (SVD)*: it exploits Singular Value Decomposition to uncover latent semantic structures; *Semantic Similarity*: it measures the semantic relatedness between terms or concepts to infer axioms or synonyms; *Stanford CoreNLP*: it uses the Stanford CoreNLP toolkit to perform natural language processing tasks, such as tokenization, parsing, and named entity recognition; *LSTM*: it employs LSTM to learn semantic patterns from textual data; *Model checking*: it applies model checking techniques to verify the consistency and correctness of ontology structures; *Clustering*: it groups similar concepts or entities into clusters to facilitate ontology generation; *Data Mapping*: it maps concepts and relationships from structured or semi-structured data sources into an ontology; *LLM-based*: it leverages LLMs to automatically extract and generate concepts and relationships from data; *Association Rules*: it discovers frequent associations and dependency rules within data to identify semantic relationships.

For the classification, we are going to use two distinct, although related, methodologies. First, we leverage pre-trained sentence transformers Nikolaev and Padó (2023) (i) to create embeddings starting from textual features. Secondly, we apply the methodology proposed by Huang et al. (2024b) (ii), in which an LLM-based clustering approach is proposed as a combination of a generation and classification task. These two classifications benefit from the

Paper	Limitation(s)
Rani et al. (2017)	Latent Discovery Analysis fails in capturing co-occurrences of topics.
Ma et al. (2021)	More databases need to be tested.
Doumanas et al. (2024)	More domains need to be explored, the choice of the LLM has a strong impact, and scalability matters are not considered.
Huang et al. (2024b)	Fully dependent on LLMs, low control, and ambiguity in the input documents.
Others	No explicit limitations.

Table 9. Limitations of semi-automatic ontology creation papers.

fact of being *agnostic* about the dimensionality of the input data.

(i) The sentence transformer model is based on *all-MiniLM-L6-v2*[**]. Following the traditional guidelines for choosing a reasonable k as number of clusters, we opted for the *silhouette score*, evaluating k from 3 to 5 and the *k-Means* as algorithm. Figure 9 shows a 2D visualisation of the cited works, with 5 as the best k.

(ii) Huang and He (2024) proposed a methodology based on the generation of labels for the instance set, and then a clustering based on the same set of labels, reformulating a clustering method as a classification one. Leveraging `GPT-3`, these four labels have been provided as output in the first-generation task, with the corresponding meanings:

- *Hybrid-expert*: Manual method involves expert creation or revision; the automatic method uses standard NLP or statistical techniques.
- *Hybrid-advanced*: Manual method involves expert creation or revision; automatic method uses advanced AI like LLMs or specific procedures.
- *Scrape-AI*: Manual method is scraping; automatic method uses advanced AI or clustering.
- *Database-AI*: Manual method is structured (e.g., relational database or ontology); automatic method uses advanced AI.

The prompt has been designed to guide the model to select a consistent and general enough set of clusters for the proposed dataset, without specifying a range of numbers of clusters. The four classes are not only the result of the `GPT` model, but have also been proposed by the models `Claude` and `Gemini`. The labels were often similar or overlapping. We report only the `GPT` labels, but we refer the reader to the repository[††]. We also tested three open-access models: `nemotron-3-ultra-550b-a55b` (Nemotron), `laguna-m.1` (Poolside), and `dolphin-mistral-24b-venice-edition` (Mistral) to better allow the reproducibility of this experiment. In this case, we preferred the labels provided by the model based on Nemotron. To provide a visual correspondence among the answers provided by the different models, we show two Sankey diagrams in Figure 11 and 12 in which we show (for paid and open-access) how many papers in a cluster are present in another cluster. These representations indicate that, for paid models, `GPT` and `Gemini` tend to agree more than `Sonnet` on the meaning of the clusters, while the free models seem less susceptible to model choice.

The designed prompt is the following one:

```
Given this dataset of semi-automatic
ontology construction approaches,
[dataset]
can you identify general clusters to
group them? Return only the names
of the clusters and their basic
characteristics?
```

Then, Table 10 shows how the LLM classified the cited works according to the labels it designed, while the full created table is available in the Appendix.

Discussion The two classifications, although distinct, do not lead to dissimilar results. In fact, more than 75% of points labelled in the same cluster for the (i) classification are also in the same cluster for the (ii) one. As shown in Figure 9, the points *Kommineni et al.* and *Qiu et al.* are between *Cluster 0* and *Cluster 4*, and they have been classified as *0*, while (ii) classified them as the equivalent of *4*. We notice a correspondence between the clusters and the labels provided by the LLM:

- 0: Hybrid-expert
- 1: Scrape-AI
- 2-3: Database-AI
- 4: Hybrid-advanced

Indeed, clusters *2* and *3* are not distinguished in the (ii) classification. This comparison supports the hypothesis that the LLM is capable of capturing the features we used for methodology comparison, capable of reproducing (or improving) a more *geometrical*-based classification based on embeddings. Finally, we provide a visual representation of the different methodologies adopted in semi-automatic approaches, as well as the frequency for every combination. In Figure 10, we show two pie charts to indicate the frequency of the manual methodologies employed (left) and the ML models (right).

Automatic Ontology Creation

Automatic ontology creation answers the need to exploit large structured knowledge to create and/or populate ontologies. This research line has a strong intersection with the NLP field, while also contributing to it by creating new scenarios and challenges. Apart from the traditional language challenges (interpretation, POS-tagging, entity disambiguation), this research also addresses consistency preservation, rule generation, and contextual reasoning, all of them while considering account performance or scalability matters. The evolution of this research follows all the major breakthroughs of Machine/Deep Learning

[**] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

[††] https://github.com/davide1797/ontology_constructions/blob/main/res/LLM_semiautomatic.txt

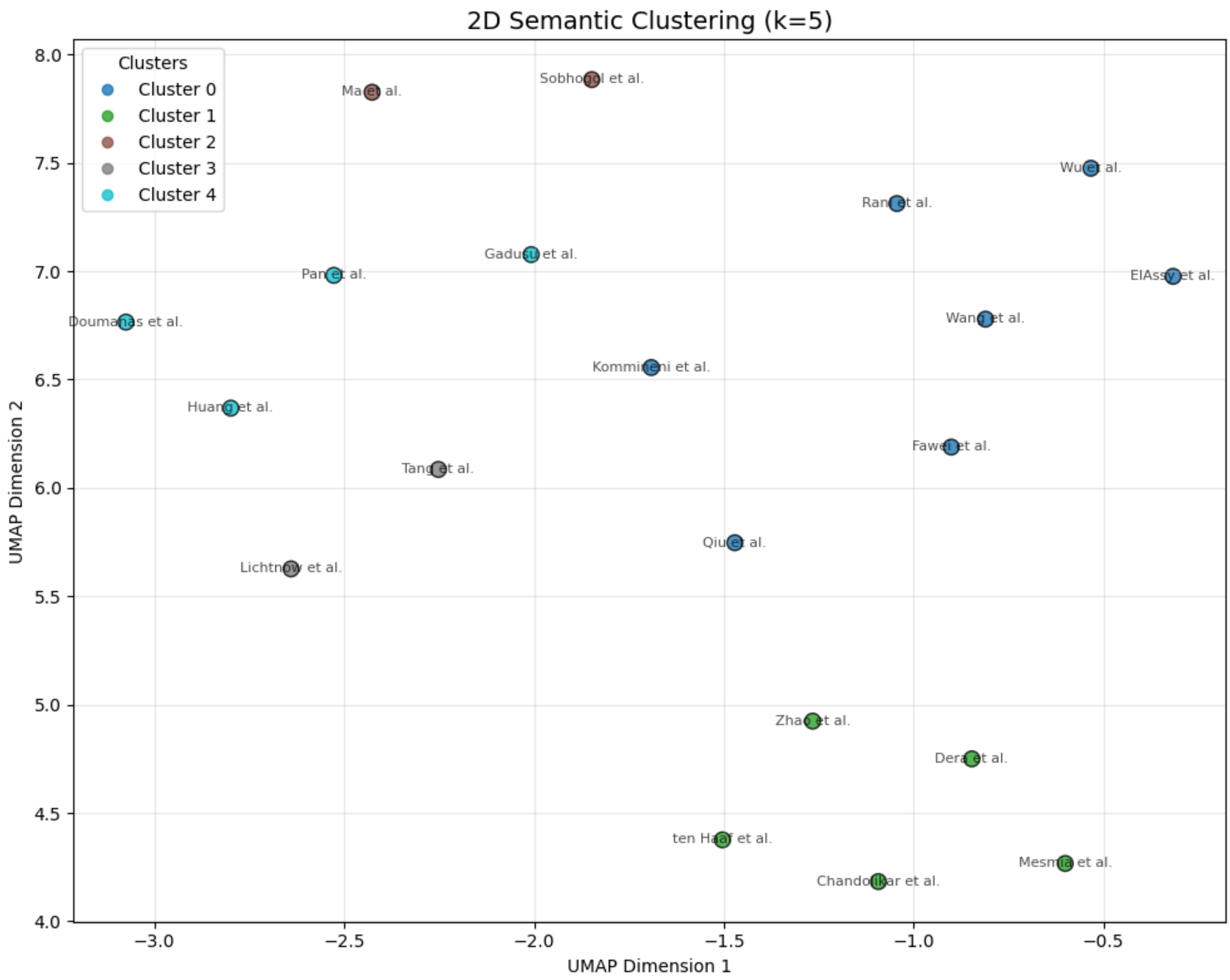
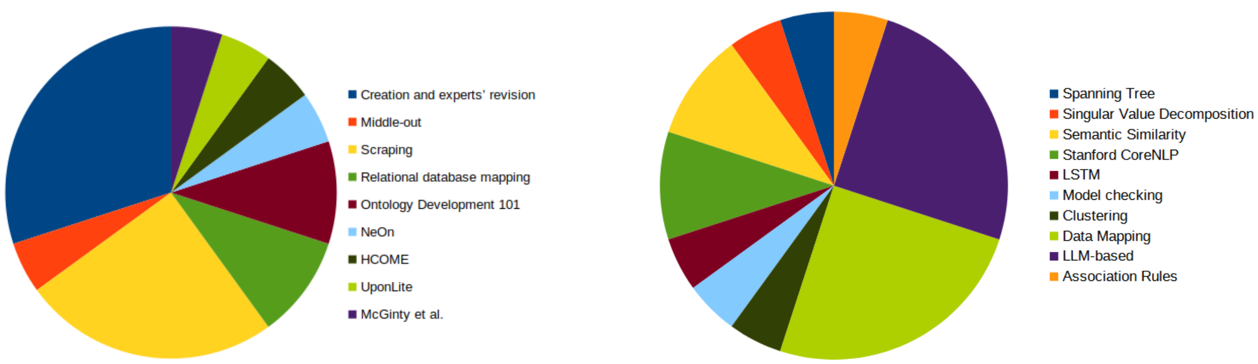


Figure 9. 2D Plot of the Classification of Semi-automatic Approaches.

LLM Label	Paper(s)
Hybrid-expert	Wang et al. (2017a); Rani et al. (2017); Fawei et al. (2019); Dera et al. (2020); Wu et al. (2025)
Hybrid-advanced	Qiu et al. (2017); ElAssy et al. (2022); Pan et al. (2023); Kommineni et al. (2024); Doumanas et al. (2024); Huang et al. (2024b); Gadusu et al. (2025)
Scrape-AI	Chandolikar et al. (2019); Zhao et al. (2021); ten Haaf et al. (2021); Mesmia and Mouhoub (2023)
Database-AI	Ma et al. (2021); Tang et al. (2023b); Sobhgol et al. (2023); Lichtnow et al. (2025)

Table 10. LLM-based classification for semi-automatic ontology approaches.



(a) Manual Methodologies in Semi-automatic Approaches.

(b) Automatic Methodologies in Semi-automatic Approaches.

Figure 10. Frequencies for manual and automatic methodologies in semi-automatic approaches.

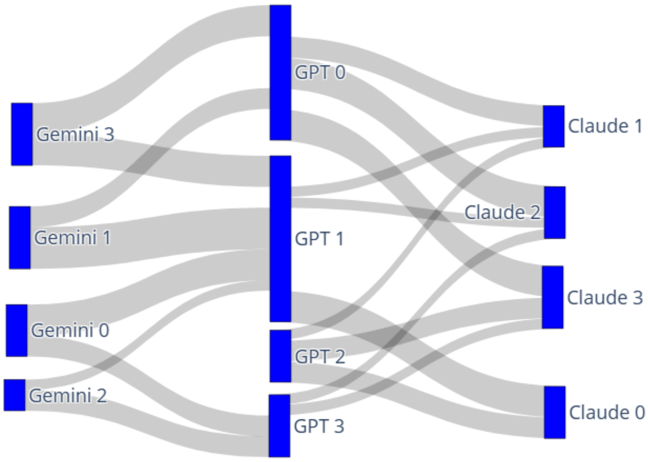


Figure 11. Sankey Diagrams for the correspondences of approaches identified by the models Gemini, GPT, and Claude.

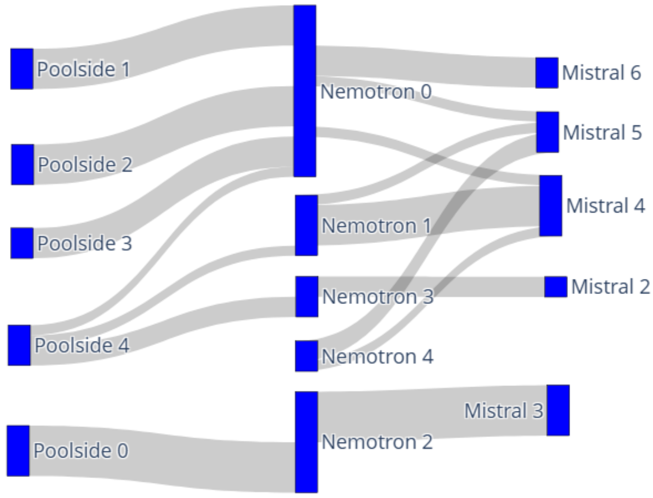


Figure 12. Sankey Diagrams for the correspondences of approaches identified by the models Poolside, Nemotron, and Mistral.

in the last few decades, given that all the most popular models have been adopted within the scope. The contribution of this section includes a classification based on two dimensions: the adopted strategies and the solved tasks. This demonstrates that technological breakthroughs have also given the possibility of tackling new tasks.

RQ$_4$: What metrics are available? For automatic ontology creation proposals, the distinguishing features are also the metrics used to make a comparison among the methodologies. Following the Machine Learning metrics, the evaluation is mostly computed as the coverage of the correct axioms (triples) automatically extracted. We report in Table 11 the list of metrics used for the different approaches. Compared to the semi-automatic approaches, this category presents more variety in tasks, which is also responsible for more metrics. We also found more metrics concerning the engineering part: apart from OOPS, we also encountered the A(ttribute)/I(nheritance)/R(elationships)/C(oncept) Richness that provide a glance at the information distribution of the ontology; for instance, highlight if the number of properties is not proportionate to the number of classes. More error functions are also available, but most papers still rely on the most common metrics.

RQ$_5$: Is there any starting methodology/dataset adopted? In some cases, extensions or improved variants of existing models are available. For automatic methodologies, datasets are crucial, and many considerations depend on them. We report in Table 12 the resources used in the papers. Here, it is much more evident that the resources are widely adopted in literature. Their size, generality, and quality allow pipelines for automatic ontology learning.

RQ$_6$: What is the context of the methodology? As for manual methodologies, some approaches may benefit from the specificity of the domain. We report in Table 13 the context (when available) of the proposed approach. Compared to the previous approaches, here more tasks related to linguistics are present, which are indeed supported by large datasets. Domain-specific contexts are also present, although less common than other ones.

RQ$_7$: What are the limitations of the methodology? Limitations may arise when the proposal is too limited to a domain or the evaluation is lacking in comparison. Limitations are shown in Table 14. As reported by the authors themselves, the lack of a human evaluation hampers the trustworthiness of the results. The output of these approaches can be fully adopted to improve some ML tasks, but they are rarely resources to be shared by practitioners. This is also one of the other major differences between this class of approaches and the others: for this one, the benefit is, in general, a new large source coming from heterogeneous data or an enrichment of existing ones. These outputs are not intended to be shared, manipulated, or reused by other researchers manually due to their size, complexity, and lack of traceability in most cases. Hence, the dimensions of size and purpose distinguish the two extremes: while the manual approaches tend to produce smaller and reusable resources, the automatic ones tend to produce bigger and trainable outputs.

RQ$_8$: How to classify the methodologies? Reviews generally propose information about the technologies used to achieve the task. In other cases, the authors preferred to distinguish the learning tasks themselves. Ontology construction can indeed refer, depending on the approach used, to axiom learning, taxonomy learning, relation extraction, etc. For this reason, we combine the two dimensions, providing a complete overview of the presented approaches, addressing both the *what* and *how* dimensions. The classification is split into two ranges of time: 2016-2023 (Figure 13), and 2024-2025 (Figure 14).

Discussion With this classification, we are able to provide partial views of the landscape, organised both by the type of task and the architecture that has been employed to address it. Along the x axis, one can observe the specific tasks that are achievable with a particular architecture, illustrating the range of applications for each model type. Conversely, along the y axis, it is evident which architectures have been utilised to tackle each individual task, providing insight into the diversity of approaches adopted in practice. It is worth noting that in the period spanning 2024 to 2025, several new tasks have emerged, reflecting the continuous evolution of

Metric	Paper(s)
Cohen's Kappa	Casteleiro et al. (2016); Albarghothi et al. (2018); Kaushik and Chatterjee (2018)
Ontological Loss and Improvement	Casteleiro et al. (2016)
Precision	Albukhitan et al. (2017); Hamano et al. (2017); Albarghothi et al. (2018); Kaushik and Chatterjee (2018); Zhao et al. (2019); Li and Chen (2020); Chen and Gu (2021); Huang et al. (2021); Lakzaei and Shamsfard (2021); Tissaoui et al. (2022); Al-Aswadi et al. (2023); Hari and Kumar (2023); He et al. (2023); Taye et al. (2023); Zhao et al. (2023); Al Subhi et al. (2024); Bakker et al. (2024); Basheer and Alkhatib (2024); Goyal et al. (2024); Huang et al. (2024a); Lo et al. (2024); Mai et al. (2024); Pisu et al. (2024); Sanaei et al. (2024); Toro et al. (2024); Ymele and Jiomekong (2024)
Recall	Albukhitan et al. (2017); Hamano et al. (2017); Albarghothi et al. (2018); Kaushik and Chatterjee (2018); Zhao et al. (2019); Li and Chen (2020); Chen and Gu (2021); Hari and Kumar (2023); Huang et al. (2021); Lakzaei and Shamsfard (2021); Tissaoui et al. (2022); Al-Aswadi et al. (2023); He et al. (2023); Taye et al. (2023); Zhao et al. (2023); Al Subhi et al. (2024); Bakker et al. (2024); Basheer and Alkhatib (2024); Dong et al. (2024); Goyal et al. (2024); Huang et al. (2024a); Lo et al. (2024); Mai et al. (2024); Pisu et al. (2024); Sanaei et al. (2024); Toro et al. (2024); Ymele and Jiomekong (2024)
Accuracy	Cahyani and Wasito (2017); Hamano et al. (2017); Kaushik and Chatterjee (2018); Petrucci et al. (2018); Zhao et al. (2019); Fumagalli et al. (2020); Tissaoui et al. (2022); Al Subhi et al. (2024); Huang et al. (2024a); Pisu et al. (2024); Bosso et al. (2025)
Coverage	Mathews and Kumar (2017)
F-Score	Albarghothi et al. (2018); Zhao et al. (2019); Li and Chen (2020); Chen and Gu (2021); Lakzaei and Shamsfard (2021); Tissaoui et al. (2022); Zhao et al. (2023); Bakker et al. (2024); Basheer and Alkhatib (2024); Goyal et al. (2024); Hari and Kumar (2023); Huang et al. (2024a); Lo et al. (2024); Mai et al. (2024); Pisu et al. (2024); Sanaei et al. (2024); Toro et al. (2024); Ymele and Jiomekong (2024)
A/I/R/C Richness	Ben Mahria et al. (2021); Chen and Gu (2021)
Mean Score Analysis	Zaouga and Rabai (2021)
Associative Probability, Perplexity	Xie et al. (2022)
p-value, t-value	Al-Aswadi et al. (2023)
Hits@K	He et al. (2023)
AUC Score	Huang et al. (2024a)
Mean Squared Error	Yin et al. (2024)
OOPS	Lippolis et al. (2025)
Non-standard metrics	Potoniec et al. (2017); Aggoune (2018); Capuano et al. (2020); Felin and Tettamanzi (2021)

Table 11. Metrics for automatic ontology construction approaches.

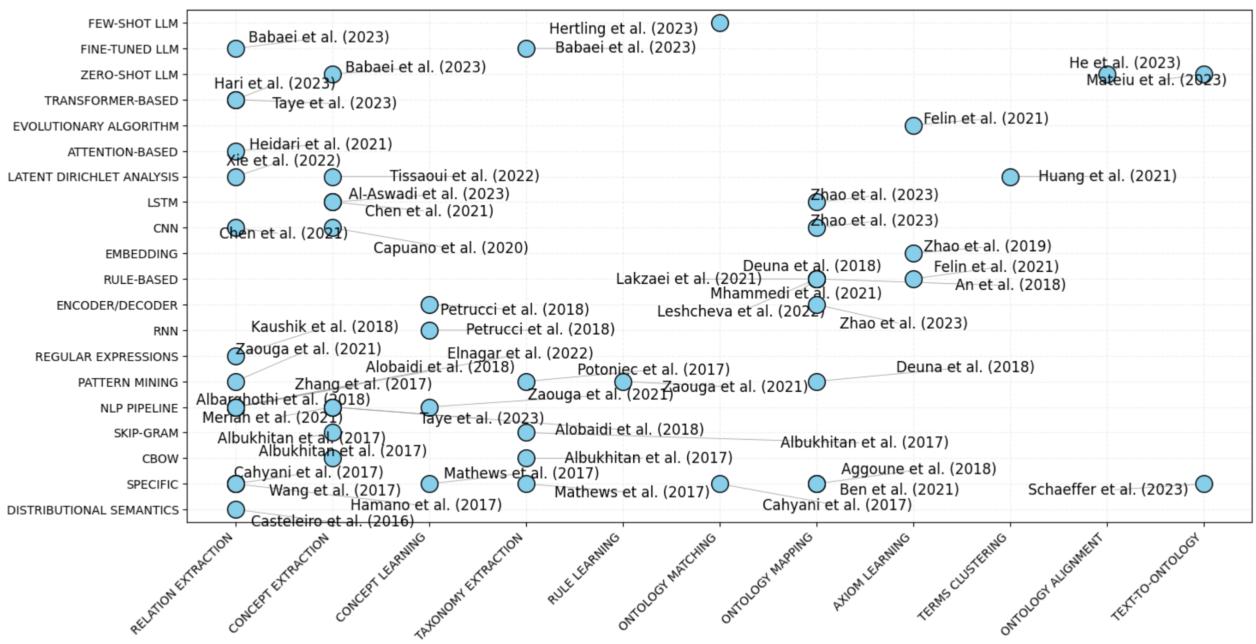


Figure 13. Classification of Automatic Strategies from 2016 to 2023.

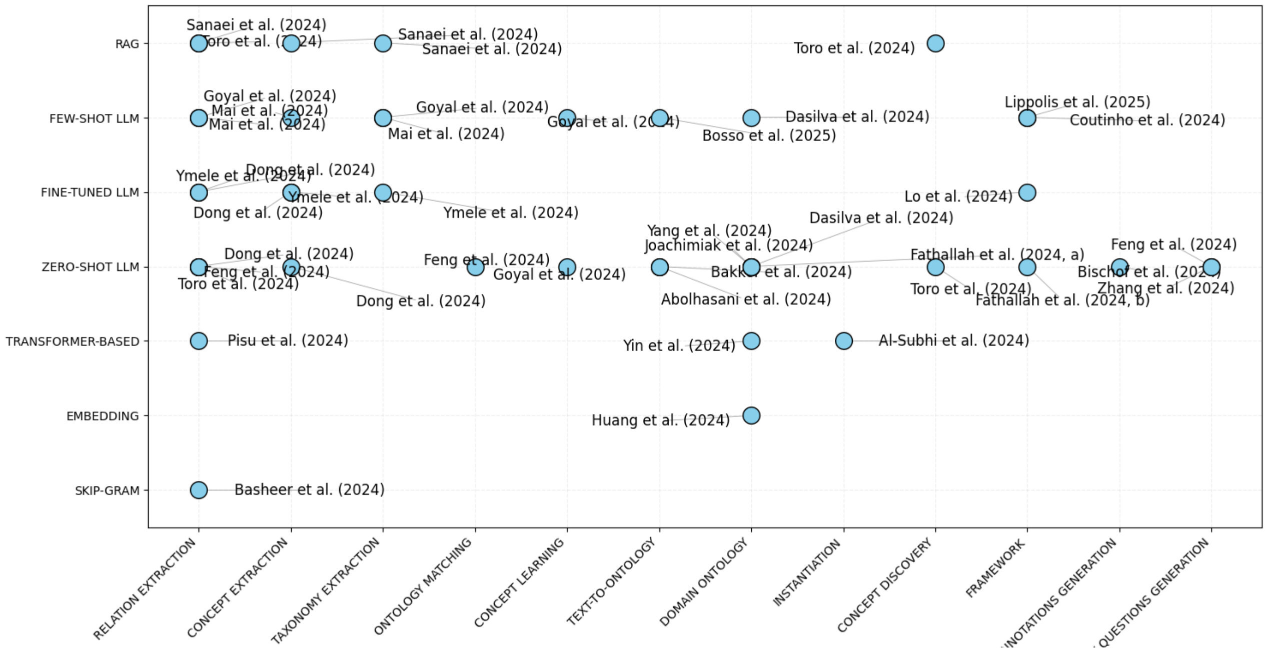


Figure 14. Classification of Automatic Strategies of 2024 and 2025.

Dataset	Paper(s)
PubMed	Casteleiro et al. (2016)
Flickr	Hamano et al. (2017)
Pizza ontology	Mathews and Kumar (2017)
DBPedia	Potoniec et al. (2017); Chen and Gu (2021); Felin and Tettamanzi (2021); Goyal et al. (2024); Sanaei et al. (2024); Ymele and Jiomekong (2024)
YAGO	Zhao et al. (2019)
CS corpus, Music corpus	Huang et al. (2021)
University, Hotel, Transportation ontology	Lakzaei and Shamsfard (2021)
NorthWind	Mhammedi et al. (2021)
New York Times, WebNLG	Hari and Kumar (2023)
Family Ontology	Mateiu and Groza (2023)
BBC, CNN, OSAc	Taye et al. (2023)
MM-S14	Dong et al. (2024)
WikiData, Wiki-NRE, SciERC, WebNLG	Feng et al. (2024)
WordNet, GeoNames, UMLS, Schema.org	Goyal et al. (2024); Sanaei et al. (2024); Ymele and Jiomekong (2024)
Gene ontology, Food ontology	Goyal et al. (2024); Sanaei et al. (2024); Ymele and Jiomekong (2024)
WikiPedia, arXiv	Lo et al. (2024)
Open English WordNet	Mai et al. (2024)
Computer Science ontology	Pisu et al. (2024)
New dataset	He et al. (2023); Joachimiak et al. (2024); Yin et al. (2024)
Non-standard datasets	Cahyani and Wasito (2017); Mathews and Kumar (2017); Wang et al. (2017b); Zhang et al. (2017); Aggoune (2018); Alobaidi et al. (2018); Kaushik and Chatterjee (2018); Petrucci et al. (2018); Capuano et al. (2020); Guimarães and de Carvalho (2020); Li and Chen (2020); Ben Mahria et al. (2021); Heidari et al. (2021); Meriah et al. (2021); Tissaoui et al. (2022); Xie et al. (2022); Al-Aswadi et al. (2023); Zhao et al. (2023); Al Subhi et al. (2024); Bakker et al. (2024); Basheer and Alkhatib (2024); Huang et al. (2024a); Toro et al. (2024); Bosso et al. (2025); Lippolis et al. (2025)

Table 12. Datasets for automatic ontology construction approaches.

the field. Within this context, it is known that large language models (LLMs) have a prominent role, often appearing in multiple variations depending on their specific training strategies, highlighting both their versatility and widespread adoption in recent research and applications.

Inter-category Discussion

The landscape of ontology construction is a continuously evolving field in which principles, ideas, and evaluations are periodically redefined, retraced or reconsidered. More precisely, technological breakthroughs like transformers, on one side, have proposed many alternatives in the field, but,

Context	Paper
Biomedicine	Casteleiro et al. (2016); Alobaidi et al. (2018)
Arabic language	Albukhitan et al. (2017); Albarghothi et al. (2018); Taye et al. (2023)
Alzheimer	Cahyani and Wasito (2017)
Language-independent, image-based	Hamano et al. (2017)
Intrusion detection	Zhang et al. (2017)
Agriculture	Kaushik and Chatterjee (2018)
Dependent on DBPedia	Capuano et al. (2020)
Portuguese language	Guimarães and de Carvalho (2020)
Security domain	Li and Chen (2020)
Relational databases	Ben Mahria et al. (2021)
Chinese language	Chen and Gu (2021)
Business	Heidari et al. (2021)
Information security	Meriah et al. (2021)
Couchbase mode	Mhammedi et al. (2021)
Project Risk Management	Zaouga and Rabai (2021)
Dependent on YAGO and DOLCE	Elnagar et al. (2022)
Constructivist Learning	Xie et al. (2022)
Topic discussion	Basheer and Alkhatib (2024)
WikiData schema	Feng et al. (2024)
Ironmaking	Huang et al. (2024a)
Artificial Intelligence	Joachimiak et al. (2024)
Threats	Bosso et al. (2025)
Unspecified context	*Others*

Table 13. Contexts of automatic ontology creation papers.

Paper	Limitation(s)
Hamano et al. (2017)	The performance depends on the images, and is problematic with concepts with multiple semantics or images with many objects.
Zhang et al. (2017)	Lack of quantitative evaluation and errors detected.
Li and Chen (2020)	For some experiments, elder approaches lead to better performances.
Basheer and Alkhatib (2024)	Much uncontrolled text mining quality of the data.
Lo et al. (2024)	Limited to taxonomical relationships
Mai et al. (2024)	Without fine-tuning, LLMs cannot adapt to specific domains.
Yin et al. (2024)	The method is time-consuming, the adopted bigram model is not satisfactory yet, and human intervention to correct inconsistencies is unavoidable.
Lippolis et al. (2025)	Lack of complete evaluation on the complete result set.
Others	No explicit limitations.

Table 14. Limitations of automatic ontology creation papers.

on the other side, new metrics are required to evaluate not only the result, but the contribution of the LLM itself. This has to be considered when deciding to follow one of the mentioned methodologies. The motivation for this contribution is the difficulty in having a clear vision of all the consequences, metrics, limitations, and technical contexts for every methodology or class of methodologies. Nonetheless, we can provide a concise summary when to choose at least one category over the others, based on the insights of the previous sections.

Manual construction is the most established research line among the three. This was due to the fact that, in the early stages of the discipline, no alternative could have been available and, still today, there are strong reasons to propose manual constructions. This line has to be generally preferred exactly when no alternatives are available; e.g., no existing data to train, no LLM trained on the concepts we desire. Although this condition may appear restrictive, in reality, it is often the case since new needs emerge frequently, for example, in nature (PFAS spread, climate change, etc) or in industrial scenarios (new data management, data migration, etc). The evaluation here relies mostly on the produced result (e.g., fairness) and on the reproducibility of the process (communication with experts, tools/diagrams, etc).

Automatic construction is quite often associated with the bottom-up approach, where some data existed before being conceived as a KG. To prefer this approach, not only should data exist, but also a pipeline of curation, alignment, and transformation should be proposed. When this line is chosen, it is expected that the data are qualitatively good enough to represent the needed concepts; hence, the evaluation is mostly based on the ML performance. The relation with experts can (and in most cases should) be present, but it is quite often limited to validation and generally not considered

a key factor in the whole process.
Semi-automatic construction represents an added value for quite well-known domains for which structured resources are missing or not suitable for a defined task. Before the LLMs, this line was conceived exactly as an intermediate layer between the other two. Currently, given the variety of tasks in which LLMs are involved, the positioning of this line is blurry. As shown above, there are some works in which LLMs are fully integrated into some manual pipelines in which they mimic the behaviour of an expert or a facilitator of formal writings (CQ generation, SPARQL queries, etc.); but there are also LLMs entering the automated part and involved in preprocessing, preparing, or summarising the data (e.g., documents) from which the graph is generated. As for the automatic part, the machine evaluation is fundamental, while some aspects related to the interaction and flow within the methodology are less stressed. In all cases, introducing LLMs raises considerations of how to evaluate their work, recognise hallucination, and manage inconsistencies. Table 15 highlights the general cases and conditions of every line of methodology.

Class	**Requirements**	**Evaluations**
Manual	Presence of human experts, awareness of the tools to design knowledge, new or unknown domain.	fairness, quality, interaction with experts, flow.
Semi-automatic	(Partial) availability of data, existing domain, fine-grained ML/LLM contribution, awareness of the ML/LLM technicalities and limitations.	ML, coverage, quality, statistical, and explanation metrics.
Automatic	Availability of large datasets, well-defined and existing domain, strong ML/LLM architecture, ML pipeline from data curation to evaluation.	ML, coverage, quality, statistical, and explanation/visualisation metrics.

Table 15. Classes of methodologies with their requirements and evaluations.

Beyond all these considerations, there are also "human limitations" to take into account. The chosen methodology is quite often governed by the methodology that the stakeholders can lead to completion and understand. This is one of the reasons why, for instance, in many interdisciplinary projects (like those involving medicine) manual graph construction is still often preferred. Time is also a fundamental dimension of the decision. Although it is impossible to say in general which methodology (or class) is faster, according to the requirements of the project, designers may prefer one solution over the other based also on this criterion.
Overall, choosing the methodology involves countless factors, and the majority of them refer to data (both input and output) and the humans responsible for producing the result. We also emphasise that the distinction of these three categories is well represented in literature by the presence of existing works D'Avanzo et al. (2008); Blomqvist (2009); Zulkipli et al. (2022) explicitly clarifying the context. We claim this has been necessary exactly to restrict the context of comparable works, given the difficulty of comparing methodologies belonging to different categories. This is also reflected by the three different evaluations provided in the manuscript.
Future developments in this research would still need to be classified according to the proposed categories, evaluated, and compared, in line with the objectives of not duplicating existing literature and keeping track of the progress.

Conclusion

In this work, we propose a two-level classification of ontology construction methodologies in the literature. The first one depicts the type of construction, which has a large impact on the needed experience, tools, or computational power. Secondly, an intra-category classification helps in understanding under which conditions the existing work can be reused, and whether some works can be grouped according to these reusability features. We claim that this work fulfils two objectives: to update the state-of-the-art in the field, especially in the areas of manual conceptualisation and LLM support, and to provide end-users with the necessary information about the methodologies they can use and the possible alternatives. Finally, this work intends to represent a current functional classification, but different facets can be explored. With the progressive injection of LLMs within manual methodologies, the distinction between the categories will likely become blurrier, and new (possibly simpler) conditions may emerge.

References

Abdelghany AS, Darwish NR and Hefni HA (2019) An agile methodology for ontology development. *International Journal of Intelligent Engineering and Systems* 12(2): 170–181.

Abolhasani MS and Pan R (2024) Leveraging llm for automated ontology extraction and knowledge graph generation. *arXiv preprint arXiv:2412.00608* .

Ackermann MR, Blömer J, Kuntze D and Sohler C (2014) Analysis of agglomerative clustering. *Algorithmica* 69(1): 184–215.

Adenowo AA and Adenowo BA (2013) Software engineering methodologies: a review of the waterfall model and object-oriented approach. *International Journal of Scientific & Engineering Research* 4(7): 427–434.

Aggoune A (2018) Automatic ontology learning from heterogeneous relational databases: Application in alimentation risks field. In: *IFIP International Conference on Computational Intelligence and Its Applications*. Springer, pp. 199–210.

Ahaggach H, Abrouk L and Lebon E (2024) Systematic mapping study of sales forecasting: Methods, trends, and future directions. *Forecasting* 6(3): 502–532.

Ahmad MN, Badr KBA, Salwana E, Zakaria NH, Tahar Z and Sattar A (2018) An ontology for the waste management domain. In: *PACIS*. p. 12.

Al-Aswadi FN, Chan HY, Gan KH et al. (2023) Enhancing relevant concepts extraction for ontology learning using domain time relevance. *Information Processing & Management* 60(1): 103140.

Al-Aswadi FN, Yong CH and Gan KH (2020) Automatic ontology construction from text: a review from shallow to deep learning trend. *The Artificial Intelligence Review* 53(6): 3901–3928.

Al Subhi SNS, Tiwari C and Mikler AR (2024) Automating hazard-specific ontology construction: Methodological advancements through ontology learning techniques from disaster-related knowledge bases. In: *2024 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*. IEEE, pp. 1–7.

Albarghothi A, Saber W and Shaalan K (2018) Automatic construction of e-government services ontology from arabic webpages. *Procedia computer science* 142: 104–113.

Albukhitan S, Helmy T and Alnazer A (2017) Arabic ontology learning using deep learning. In: *Proceedings of the international conference on web intelligence*. pp. 1138–1142.

Alobaidi M, Malik KM and Sabra S (2018) Linked open data-based framework for automatic biomedical ontology generation. *BMC bioinformatics* 19: 1–13.

An J and Park YB (2018) Methodology for automatic ontology generation using database schema information. *Mobile Information Systems* 2018(1): 1359174.

Angles R (2018) The property graph database model. In: *AMW*.

Angles R, Arenas M, Barceló P, Hogan A, Reutter J and Vrgoč D (2017) Foundations of modern query languages for graph databases. *ACM Computing Surveys (CSUR)* 50(5): 1–40.

Armary P, El-Vaigh CB, Narsis OL and Nicolle C (2025) Ontology learning towards expressiveness: A survey. *Computer Science Review* 56: 100693.

Arrar D, Kamel N and Lakhfif A (2024) A comprehensive survey of link prediction methods: D. arrar et al. *The journal of supercomputing* 80(3): 3902–3942.

Asim MN, Wasim M, Khan MUG, Mahmood W and Abbasi HM (2018) A survey of ontology learning techniques and applications. *Database* 2018: bay101.

Babaei Giglou H, D'Souza J and Auer S (2023) Llms4ol: Large language models for ontology learning. In: *International Semantic Web Conference*. Springer, pp. 408–427.

Badr KBA, Badr ABA and Ahmad MN (2013) Phases in ontology building methodologies: A recent review. *Ontology-Based Applications for Enterprise Systems and Knowledge Management* : 100–123.

Bakker RM, Di Scala DL and de Boer M (2024) Ontology learning from text: an analysis on llm performance. In: *Proceedings of the 3rd NLP4KGC International Workshop on Natural Language Processing for Knowledge Graph Creation, colocated with Semantics*. pp. 17–19.

Basheer R and Alkhatib B (2024) Darkonto: An ontology construction approach for dark web community discussions through topic modeling and ontology learning. *Human Behavior and Emerging Technologies* 2024(1): 7914028.

Bautista-Zambrana MR (2015) Methodologies to build ontologies for terminological purposes. *Procedia-Social and Behavioral Sciences* 173: 264–269.

Ben Mahria B, Chaker I and Zahi A (2021) A novel approach for learning ontology from relational database: from the construction to the evaluation. *Journal of Big Data* 8(1): 25.

Benjamins VR and Fensel D (1998) The ontological engineering initiative (ka) 2. In: *Formal Ontology in Information Systems*. Citeseer, pp. 287–301.

Berners-Lee T et al. (1998) Semantic web road map.

Bischof S, Filtz E, Parreira JX and Steyskal S (2024) Llm-based guided generation of ontology term definitions. In: *European Semantic Web Conference*. Springer, pp. 133–137.

Blomqvist E (2009) *Semi-automatic ontology construction based on patterns*. Linkopings Universitet (Sweden).

Borst P, Akkermans J, Pos A and Top J (1995) The physsys ontology for physical systems. In: *Working Papers of the Ninth International Workshop on Qualitative Reasoning QR*, volume 95. pp. 11–21.

Bosso F, Ruberto S et al. (2025) Ontology learning for hybrid threats .

Cahyani DE and Wasito I (2017) Automatic ontology construction using text corpora and ontology design patterns (odps) in alzheimer's disease. *Jurnal Ilmu Komputer dan Informasi* 10(2): 59–66.

Capuano A, Rinaldi AM and Russo C (2020) An ontology-driven multimedia focused crawler based on linked open data and deep learning techniques. *Multimedia Tools and Applications* 79(11): 7577–7598.

Casteleiro MA, Prieto MJF, Demetriou G, Maroto N, Read WJ, Maseda-Fernandez D, Des Diz JJ, Nenadic G, Keane JA and Stevens R (2016) Ontology learning with deep learning: a case study on patient safety using pubmed. In: *SWAT4LS*.

Chandolikar N, Shilaskar S, Peddawad D and Bhosale S (2019) Semi-automated ontology building using deep learning to provide domain-specific knowledge search in the marathi language. In: *2019 International conference on applied machine learning (ICAML)*. IEEE, pp. 108–113.

Chang Y, Wang X, Wang J, Wu Y, Yang L, Zhu K, Chen H, Yi X, Wang C, Wang Y et al. (2024) A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology* 15(3): 1–45.

Chen J and Gu J (2021) Adol: a novel framework for automatic domain ontology learning. *The Journal of Supercomputing* 77(1): 152–169.

Chen X, Jia S and Xiang Y (2020) A review: Knowledge reasoning over knowledge graph. *Expert systems with applications* 141: 112948.

Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Moschitti A, Pang B and Daelemans W (eds.) *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 1724–1734. DOI:10.3115/v1/D14-1179.

Chowdhary K and Chowdhary K (2020) Natural language processing. *Fundamentals of artificial intelligence* : 603–649.

Chowdhery A et al. (2022) Scaling language modeling with pathways.

Coutinho ML (2024) Leveraging llms in text-based ontology driven conceptual modeling.

Cristani M and Cuel R (2005) A survey on ontology creation methodologies. *International Journal on Semantic Web and Information Systems (IJSWIS)* 1(2): 49–69.

da Silva LMV, Kocher A, Gehlhoff F and Fay A (2024) On the use of large language models to generate capability ontologies. In: *2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, pp. 1–8.

Dahlem N and Hahn A (2009) User-friendly ontology creation methodologies-a survey. *AMCIS 2009 Proceedings* : 117.

Daponte V and Falquet G (2018) Une ontologie pour la formalisation et la visualisation des connaissances scientifiques. In: *29es Journées Francophones d'Ingénierie des Connaissances, IC 2018*. pp. 129–136.

De Nicola A and Missikoff M (2016) A lightweight methodology for rapid ontology engineering. *Communications of the ACM* 59(3): 79–86.

De Nicola A, Missikoff M and Navigli R (2005) A proposal for a unified process for ontology building: Upon. In: *International conference on database and expert systems applications*. Springer, pp. 655–664.

De Uña D, Rümmele N, Gange G, Schachte P and Stuckey PJ (2018) Machine learning and constraint programming for relational-to-ontology schema mapping. In: *International Joint Conference on Artificial Intelligence 2018*. Association for the Advancement of Artificial Intelligence (AAAI), pp. 1277–1283.

Debruyne C and Meersman R (2012) Gospl: A method and tool for fact-oriented hybrid ontology engineering. In: *East European Conference on Advances in Databases and Information Systems*. Springer, pp. 153–166.

Dera E, Frasincar F, Schouten K and Zhuang L (2020) Sasobus: Semi-automatic sentiment domain ontology building using synsets. In: *European Semantic Web Conference*. Springer, pp. 105–120.

Dong H, Chen J, He Y, Gao Y and Horrocks I (2024) A language model based framework for new concept placement in ontologies. In: *European Semantic Web Conference*. Springer, pp. 79–99.

Doumanas D, Bouchouras G, Soularidis A, Kotis K and Vouros G (2024) From human-to llm-centered collaborative ontology engineering. *Applied Ontology* 19(4): 334–367.

Dutta B, Chatterjee U and Madalli DP (2015) Yamo: yet another methodology for large-scale faceted ontology construction. *Journal of Knowledge Management* 19(1): 6–24.

D'Avanzo E, Lieto A and Kuflik T (2008) *Manually vs semiautomatic domain specific ontology building*. PhD Thesis, University of Salerno Baronissi, Italy.

ElAssy O, de Vendt R, Dalpiaz F and Brinkkemper S (2022) A semi-automated method for domain-specific ontology creation from medical guidelines. In: *International Conference on Business Process Modeling, Development and Support*. Springer, pp. 295–309.

Elnagar S, Yoon V and Thomas MA (2022) An automatic ontology generation framework with an organizational perspective. *arXiv preprint arXiv:2201.05910* .

Erickson BJ and Kitamura F (2021) Magician's corner: 9. performance metrics for machine learning models.

Fathallah N, Das A, Giorgis SD, Poltronieri A, Haase P and Kovriguina L (2024a) Neon-gpt: a large language model-powered pipeline for ontology learning. In: *European Semantic Web Conference*. Springer, pp. 36–50.

Fathallah N, Staab S and Algergawy A (2024b) Llms4life: Large language models for ontology learning in life sciences. *arXiv preprint arXiv:2412.02035* .

Fawei B, Pan JZ, Kollingbaum M and Wyner AZ (2018) A methodology for a criminal law and procedure ontology for legal question answering. In: *Semantic Technology: 8th Joint International Conference, JIST 2018, Awaji, Japan, November 26–28, 2018, Proceedings 8*. Springer, pp. 198–214.

Fawei B, Pan JZ, Kollingbaum M and Wyner AZ (2019) A semi-automated ontology construction for legal question answering. *New Generation Computing* 37(4): 453–478.

Felin R and Tettamanzi AG (2021) Using grammar-based genetic programming for mining subsumption axioms involving complex class expressions. In: *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. pp. 234–240.

Feng X, Wu X and Meng H (2024) Ontology-grounded automatic knowledge graph construction by llm under wikidata schema. *arXiv preprint arXiv:2412.20942* .

Fernández-López M and Gómez-Pérez A (2002) Overview and analysis of methodologies for building ontologies. *The knowledge engineering review* 17(2): 129–156.

Fernández-López M, Gómez-Pérez A and Juristo N (1997) Methontology: from ontological art towards ontological engineering .

Fox MS (1992) The tove project towards a common-sense model of the enterprise. In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, pp. 25–34.

Fumagalli M, Bella G, Conti S and Giunchiglia F (2020) Ontology-driven cross-domain transfer learning. In: *Formal Ontology in Information Systems*. IOS Press, pp. 249–263.

Gadusu SR, Callahan L, Lababidi S, Nishtala A, Healey S and McGinty H (2025) Generating reliable adverse event profiles for health through automated integrated data (graph-aid): A semi-automated ontology building approach. *arXiv preprint arXiv:2506.20851* .

Gahleitner E, Behrendt W, Palkoska J and Weippl E (2005) On cooperatively creating dynamic ontologies. In: *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*. pp. 208–210.

Gangemi A, Guarino N, Masolo C and Oltramari A (2003) Sweetening wordnet with dolce. *AI magazine* 24(3): 13–13.

Gangemi A, Steve G and Giacomelli F (1996) Onions: An ontological methodology for taxonomic knowledge integration. In: *ECAI-96 Workshop on Ontological Engineering, Budapest*, volume 95.

Genesereth MR and Nilsson NJ (2012) *Logical foundations of artificial intelligence*. Morgan Kaufmann.

Glimm B, Horrocks I, Motik B, Stoilos G and Wang Z (2014) Hermit: an owl 2 reasoner. *Journal of automated reasoning* 53(3): 245–269.

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53(3-4): 325–338.

Goyal PK, Singh S and Tiwary US (2024) silp_nlp at llms4ol 2024 tasks a, b, and c: Ontology learning through prompts with llms. In: *Open Conference Proceedings*, volume 4. pp. 31–38.

Gruber T (1993) What is an ontology .

Gruninger M (1995) Methodology for the design and evaluation of ontologies. In: *Proc. IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing*.

Guimarães NC and de Carvalho CL (2020) A modular framework for ontology learning from text in portuguese. *Multi-Science Journal* 3(3): 37–42.

Hamano S, Ogawa T and Haseyama M (2017) A language-independent ontology construction method using tagged images in folksonomy. *IEEE Access* 6: 2930–2942.

Hari A and Kumar P (2023) Wsd based ontology learning from unstructured text using transformer. *Procedia Computer Science* 218: 367–374.

Hastie T, Tibshirani R, Friedman JH and Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

He Y, Chen J, Dong H and Horrocks I (2023) Exploring large language models for ontology alignment. *arXiv preprint arXiv:2309.07172* .

Heidari M, Zad S, Berlin B and Rafatirad S (2021) Ontology creation model based on attention mechanism for a specific business domain. In: *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, pp. 1–5.

Hert M, Reif G and Gall HC (2011) A comparison of rdb-to-rdf mapping languages. In: *Proceedings of the 7th international conference on semantic systems*. pp. 25–32.

Hertling S and Paulheim H (2023) Olala: Ontology matching with large language models. In: *Proceedings of the 12th knowledge capture conference 2023*. pp. 131–139.

Hogan A, Blomqvist E, Cochez M, d'Amato C, Melo GD, Gutierrez C, Kirrane S, Gayo JEL, Navigli R, Neumaier S et al. (2021) Knowledge graphs. *ACM Computing Surveys (Csur)* 54(4): 1–37.

Hristozova M and Sterling L (2002) An extreme method for developing lightweight ontologies. In: *Workshop on Ontologies in Agent Systems, 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems,(Bologna, Italy, 2002)*.

Huang C and He G (2024) Text clustering as classification with llms. *arXiv preprint arXiv:2410.00927* .

Huang H, Harzallah M, Guillet F and Xu Z (2021) Core-concept-seeded lda for ontology learning. *Procedia computer science* 192: 222–231.

Huang X, Yang C, Zhang Y, Lou S, Kong L and Zhou H (2024a) Ontology guided multi-level knowledge graph construction and its applications in blast furnace ironmaking process. *Advanced Engineering Informatics* 62: 102927.

Huang Y, Karabulut E and Degeler V (2024b) Large language model for ontology learning in drinking water distribution network domain. In: *EKAW (Companion)*.

Huang Z, Yang W, Li X and Bai Q (2026) Multi-source knowledge graph construction through llm-assisted incremental fusion. *Intelligent Systems with Applications* 31: 200674. DOI:https://doi.org/10.1016/j.iswa.2026.200674. URL https://www.sciencedirect.com/science/article/pii/S2667305326000499.

Iqbal R, Murad MAA, Mustapha A, Sharef NM et al. (2013) An analysis of ontology engineering methodologies: A literature review. *Research journal of applied sciences, engineering and technology* 6(16): 2993–3000.

Jacksi K (2019) Design and implementation of e-campus ontology with a hybrid software engineering methodology. *Science Journal of University of Zakho* 7(3): 95–100.

Jarrar M and Meersman R (2009) Ontology engineering–the dogma approach. *Advances in Web Semantics I* : 7–34.

Jiang H, Jin H, Zhang M, Li X, Lyu K, Deng W and Wang D (2026) Dressinggraphrag: A domain-specific research assistant integrating graphrag and llms for wound dressing research. *Chinese Chemical Letters* : 112632DOI:https://doi.org/10.1016/j.cclet.2026.112632. URL https://www.sciencedirect.com/science/article/pii/S1001841726002755.

Joachimiak MP, Miller MA, Caufield JH, Ly R, Harris NL, Tritt A, Mungall CJ and Bouchard KE (2024) The artificial intelligence ontology: Llm-assisted construction of ai concept hierarchies. *Applied Ontology* 19(4): 408–418.

John S, Shah N and Stewart C (2018) Towards a software centric approach for ontology development: Novel methodology and its application. In: *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*. IEEE, pp. 139–146.

Jones DM, Bench-Capon TJM and Visser PRS (2007) Methodologies for ontology development. URL https://api.semanticscholar.org/CorpusID:8263022.

Jordan MI and Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* 349(6245): 255–260.

Kaushik N and Chatterjee N (2018) Automatic relationship extraction from agricultural text for ontology construction. *Information processing in agriculture* 5(1): 60–73.

Khadir AC, Aliane H and Guessoum A (2021) Ontology learning: Grand tour and challenges. *Computer Science Review* 39: 100339.

Kohlhase M (2000) Omdoc: An infrastructure for openmath content dictionary information. *ACM SIGSAM Bulletin* 34(2): 43–48.

Kommineni VK, König-Ries B and Samuel S (2024) From human experts to machines: An llm supported approach to ontology and knowledge graph construction. *arXiv preprint arXiv:2403.08345* .

Kotis K and Vouros GA (2006) Human-centered ontology engineering: The hcome methodology. *Knowledge and Information Systems* 10: 109–131.

Lakzaei B and Shamsfard M (2021) Ontology learning from relational databases. *Information Sciences* 577: 280–297.

LeCun Y, Bengio Y and Hinton G (2015) Deep learning. *nature* 521(7553): 436–444.

Lenat DB, Prakash M and Shepherd M (1985) Cyc: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI magazine* 6(4): 65–65.

Leshcheva I and Begler A (2022) A method of semi-automated ontology population from multiple semi-structured data sources. *Journal of information science* 48(2): 223–236.

Li J and Alian S (2018) Design and development of a biocultural ontology for personalized diabetes self-management of american indians. In: *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, pp. 1–7.

Li J, Garijo D and Poveda-Villalón M (2026) Large language models for ontology engineering: a systematic literature review. *Semantic Web* 17(4): 22104968261465514.

Li T and Chen Z (2020) An ontology-based learning approach for automatically classifying security requirements. *Journal of Systems and Software* 165: 110566.

Li Z, Yang MC and Ramani K (2009) A methodology for engineering ontology acquisition and validation. *AI EDAM* 23(1): 37–51.

Lichtnow D, Fleischmann AMP, do Nascimento LV, Machado GM and de Oliveira JPM (2025) Pipeline for ontology construction using a large language model: A smart campus use case. In: *Proceedings of the 27th International Conference on Enterprise Information Systems, ICEIS*. pp. 4–6.

Lippolis AS, Saeedizade MJ, Keskisärkkä R, Zuppiroli S, Ceriani M, Gangemi A, Blomqvist E and Nuzzolese AG (2025) Ontology generation using large language models. In: *European Semantic Web Conference*. Springer, pp. 321–341.

Lo A, Jiang AQ, Li W and Jamnik M (2024) End-to-end ontology learning with large language models. *Advances in Neural Information Processing Systems* 37: 87184–87225.

Lourdusamy R and John A (2018) A review on metrics for ontology evaluation. In: *2018 2nd International conference on inventive systems and control (ICISC)*. IEEE, pp. 1415–1421.

Ma C, Molnár B and Benczúr A (2021) A semi-automatic semantic consistency-checking method for learning ontology from relational database. *Information* 12(5): 188.

Mahmood K, Mokhtar R, Raza MA, Noraziah A and Alkazemi B (2022) Ecological and confined domain ontology construction scheme using concept clustering for knowledge management. *Applied sciences* 13(1): 32.

Mai HT, Chu CX and Paulheim H (2024) Do llms really adapt to domains? an ontology learning perspective. In: *International Semantic Web Conference*. Springer, pp. 126–143.

Martínez V, Berzal F and Cubero JC (2016) A survey of link prediction in complex networks. *ACM computing surveys (CSUR)* 49(4): 1–33.

Mateiu P and Groza A (2023) Ontology engineering with large language models. In: *2023 25th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. IEEE, pp. 226–229.

Mathews KA and Kumar PS (2017) Extracting ontological knowledge from textual descriptions. *arXiv preprint arXiv:1709.08448* .

McGinty HK (2018) *KNowledge Acquisition and Representation Methodology (KNARM) and its applications*. PhD Thesis, University of Miami.

Meriah I, Rabai LBA and Khedri R (2021) Towards an automatic approach to the design of a generic ontology for information security. In: *2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS)*. IEEE, pp. 1–8.

Mesmia FB and Mouhoub M (2023) Semi-automatic building and learning of a multilingual ontology. *ACM Transactions on Asian and Low-Resource Language Information Processing* 22(11): 1–19.

Mhammedi S, El Massari H and Gherabi N (2021) Cb2onto: Owl ontology learning approach from couchbase. In: *Intelligent Systems in Big Data, Semantic Web and Machine Learning*. Springer, pp. 95–110.

Michel F, Montagnat J and Zucker CF (2014) *A survey of RDB to RDF translation approaches and tools*. PhD Thesis, I3S.

Moore P and Van Pham H (2015) On context and the open world assumption. In: *2015 IEEE 29th international conference on advanced information networking and applications workshops*. IEEE, pp. 387–392.

Murphy KP (2012) *Machine learning: a probabilistic perspective*. MIT press.

Nabi SI and Asif Z (2015) Grounded ontology methodology - illustrating the seed ontology creation. In: *UK Academy for Information Systems Annual Conference*. URL https://api.semanticscholar.org/CorpusID:19876964.

Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, Akhtar N, Barnes N and Mian A (2023) A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology* .

Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, Akhtar N, Barnes N and Mian A (2025) A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology* 16(5): 1–72.

Nefzi H, Farah M, Farah IR and Solaiman B (2014) A critical analysis of lifecycles and methods for ontology construction and evaluation. In: *2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, pp. 48–53.

Nikolaev D and Padó S (2023) Representation biases in sentence transformers. *arXiv preprint arXiv:2301.13039* .

Noy NF, McGuinness DL et al. (2001) Ontology development 101: A guide to creating your first ontology.

Olivares-Alarcos A, Foix S, Borgo S and Alenyà G (2022) Ocra–an ontology for collaborative robotics and adaptation. *Computers in Industry* 138: 103627.

O'Hara T, Mahesh K and Nirenburg S (1998) Lexical acquisition with wordnet and the mikrokosmos ontology. In: *Usage of WordNet in Natural Language Processing Systems*.

Pan Z, Gao Y, Ponci F and Monti A (2023) Semi-automatic ontology development framework for building energy data management. *IEEE access* 11: 111991–112003.

Peraketh B, Menzel C, Mayer RJ, Fillion F, Futrell MT, DeWitte P et al. (1994) Ontology capture method (idef5). *Knowledge*

Based Systems, Incorporated Technical report .

Peroni S (2016) Samod: an agile methodology for the development of ontologies. In: *Proceedings of the 13th OWL: Experiences and Directions Workshop and 5th OWL reasoner evaluation workshop (OWLED-ORE 2016)*. pp. 1–14.

Petersen K, Feldt R, Mujtaba S and Mattsson M (2008) Systematic mapping studies in software engineering. In: *12th international conference on evaluation and assessment in software engineering (EASE)*. BCS Learning & Development.

Petrucci G, Rospocher M and Ghidini C (2018) Expressive ontology learning as neural machine translation. *Journal of Web Semantics* 52: 66–82.

Pinto HS, Staab S and Tempich C (2004) Diligent: Towards a fine-grained methodology for distributed, loosely-controlled and evolving enginering of ontologies. In: *ECAI*, volume 16. p. 393.

Pisu A, Pompianu L, Salatino A, Osborne F, Riboni D, Motta E, Reforgiato Recupero D et al. (2024) Leveraging language models for generating ontologies of research topics. In: *CEUR WORKSHOP PROCEEDINGS*, volume 3747. CEUR-WS, p. 11.

Potoniec J, Jakubowski P and Ławrynowicz A (2017) Swift linked data miner: Mining owl 2 el class expressions directly from online rdf datasets. *Journal of Web Semantics* 46: 31–50.

Poveda-Villalón M, Fernández-Izquierdo A, Fernández-López M and García-Castro R (2022) Lot: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence* 111: 104755.

Poveda-Villalón M, Suárez-Figueroa MC and Gómez-Pérez A (2012) Validating ontologies with oops! In: *International conference on knowledge engineering and knowledge management*. Springer, pp. 267–281.

Qiu Y, Cheng L and Alghazzawi D (2017) Towards a semi-automatic method for building chinese tax domain ontology. In: *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, pp. 2530–2539.

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I et al. (2019) Language models are unsupervised multitask learners. *OpenAI blog* 1(8): 9.

Rani M, Dhar AK and Vyas OP (2017) Semi-automatic terminology ontology learning based on topic modeling. *Engineering Applications of Artificial Intelligence* 63: 108–125.

Reiter R (1988) Nonmonotonic reasoning. In: *Exploring artificial intelligence*. Elsevier, pp. 439–481.

Rokach L and Maimon O (2005) Clustering methods. In: *Data mining and knowledge discovery handbook*. Springer, pp. 321–352.

Saad A and Shaharin S (2016) The methodology for ontology development in lesson plan domain. *Methodology* 7(4).

Sabou M, Wroe C, Goble C and Mishne G (2005) Learning domain ontologies for web service descriptions: an experiment in bioinformatics. In: *Proceedings of the 14th international conference on World Wide Web*. pp. 190–198.

Sanaei M, Azizi F and Giglou HB (2024) Phoenixes at llms4ol 2024 tasks a, b, and c: Retrieval augmented generation for ontology learning. In: *Open Conference Proceedings*, volume 4. pp. 39–47.

Sandkuhl K et al. (2005) Towards a methodology for ontology development in small and medium-sized enterprises .

Sattar A, Surin ESM, Ahmad MN, Ahmad M and Mahmood AK (2020) Comparative analysis of methodologies for domain ontology development: A systematic review. *International Journal of Advanced Computer Science and Applications* 11(5).

Schaeffer M, Sesboüé M, Kotowicz JP, Delestre N and Zanni-Merk C (2023) Olaf: An ontology learning applied framework. *Procedia Computer Science* 225: 2106–2115.

Sesen MB, Au Yeung J and Asgari E (2025) Development and validation of retrieval augmented generation (rag) and graphrag for complex clinical cases. *medRxiv* : 2025–11.

Shahapure KR and Nicholas C (2020) Cluster quality analysis using silhouette score. In: *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*. IEEE, pp. 747–748.

Sharma S and Jain S (2024) Covido: an ontology for covid-19 metadata. *The Journal of Supercomputing* 80(1): 1238–1267.

Shofi IM and Budiardjo EK (2012) Addressing owl ontology for goal consistency checking. In: *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*. pp. 336–341.

Shukla NK, Prabhakar P, Thangaraj S, Singh S, Sun W, Venkatesan CP and Krishnamurthy V (2025) Graphrag analysis for financial narrative summarization and a framework for optimizing domain adaptation. In: *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*. pp. 23–34.

Sirin E, Parsia B, Grau BC, Kalyanpur A and Katz Y (2007) Pellet: A practical owl-dl reasoner. *Journal of Web Semantics* 5(2): 51–53.

Smith B (2012) Ontology .

Sobhgol SS, Thron M and Persico G (2023) Towards semi-automatic approach of building an ontology: A case study on material handling data. In: *KEOD*. pp. 248–254.

Studer R, Benjamins VR and Fensel D (1998) Knowledge engineering: Principles and methods. *Data & knowledge engineering* 25(1-2): 161–197.

Suárez-Figueroa MC, Gómez-Pérez A and Fernández-López M (2011) The neon methodology for ontology engineering. In: *Ontology engineering in a networked world*. Springer, pp. 9–34.

Sugiyama M (2015) *Introduction to statistical machine learning*. Morgan Kaufmann.

Sure Y, Akkermans H, Broekstra J, Davies J, Ding Y, Duke A, Engels R, Fensel D, Horrocks I, Iosif V et al. (2003) On-to-knowledge: semantic web-enabled knowledge management. In: *Web Intelligence*. Springer, pp. 277–300.

Susmaga R (2004) Confusion matrix visualization. In: *Intelligent information processing and web mining: proceedings of the international IIS: IIPWM '04 conference held in zakopane, Poland, may 17–20, 2004*. Springer, pp. 107–116.

Swartout B, Patil R, Knight K and Russ T (1996) Toward distributed use of large-scale ontologies. In: *Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*, volume 138. p. 25.

Tailhardat L, Chabot Y and Troncy R (2024) Noria-o: an ontology for anomaly detection and incident management in ict systems. In: *European Semantic Web Conference*. Springer, pp. 21–39.

Tan PN, Steinbach M and Kumar V (2016) *Introduction to data mining*. Pearson Education India.

Tang X, Feng Z, Xiao Y, Wang M, Ye T, Zhou Y, Meng J, Zhang B and Zhang D (2023a) Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development. *Geoscience Frontiers* 14(5): 101426.

Tang Y, Da Costa AAB, Zhang X, Patrick I, Khastgir S and Jennings P (2023b) Domain knowledge distillation from large language model: An empirical study in the autonomous driving domain. In: *2023 IEEE 26th international conference on intelligent transportation systems (ITSC)*. IEEE, pp. 3893–3900.

Taye MM (2010) Understanding semantic web and ontologies: Theory and applications. *arXiv preprint arXiv:1006.4567* .

Taye MM, Abulail R and Al-Oudat M (2023) An ontology learning framework for unstructured arabic text. In: *2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*. IEEE, pp. 1–12.

ten Haaf F, Claassen C, Eschauzier R, Tjan J, Buijs D, Frasincar F and Schouten K (2021) Web-soba: Word embeddings-based semi-automatic ontology building for aspect-based sentiment classification. In: *European Semantic Web Conference*. Springer, pp. 340–355.

Thesing T, Feldmann C and Burchardt M (2021) Agile versus waterfall project management: decision model for selecting the appropriate approach to a project. *Procedia Computer Science* 181: 746–756.

Tissaoui A, Sassi S, Chbeir R and Mechergui A (2022) A top-down enriching approach for ontology learning from text. *Concurrency and Computation: Practice and Experience* 34(19): e7036.

Toro S, Anagnostopoulos AV, Bello SM, Blumberg K, Cameron R, Carmody L, Diehl AD, Dooley DM, Duncan WD, Fey P et al. (2024) Dynamic retrieval augmented generation of ontologies using artificial intelligence (dragon-ai). *Journal of Biomedical Semantics* 15(1): 19.

Touvron H, Martin L, Stone K, Albert P, Almahairi A, Babaei Y, Bashlykov N, Batra S, Bhargava P, Bhosale S et al. (2023) Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* .

Uschold M and King M (1995) *Towards a methodology for building ontologies*. Citeseer.

Van Der Maaten L, Postma EO, Van Den Herik HJ et al. (2009) Dimensionality reduction: A comparative review. *Journal of machine learning research* 10(66-71): 13.

van der Vet PE, Speel PH and Mars NJ (1994) The plinius ontology of ceramic materials. In: *Eleventh European Conference on Artificial Intelligence (ECAI'94) Workshop on Comparison of Implemented Ontologies*. pp. 8–12.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł and Polosukhin I (2017) Attention is all you need. *Advances in neural information processing systems* 30.

Wang J, Flanagan B and Ogata H (2017a) Semi-automatic construction of ontology based on data mining technique. In: *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, pp. 511–515.

Wang Y, Valipour M, Zatarain OD, Gavrilova ML, Hussain A, Howard N and Patel S (2017b) Formal ontology generation by deep machine learning. In: *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*. IEEE, pp. 6–15.

Wielinga BJ, Schreiber AT and Breuker JA (1992) Kads: A modelling approach to knowledge engineering. *Knowledge acquisition* 4(1): 5–53.

Wold S, Esbensen K and Geladi P (1987) Principal component analysis. *Chemometrics and intelligent laboratory systems* 2(1-3): 37–52.

Wu Z, Liu M and Ma G (2025) A semi-automatic ontology development framework for knowledge transformation of construction safety requirements. *Buildings* 15(4): 569.

Xie Z, Ren L, Zhan Q and Liu Y (2022) A constructivist ontology relation learning method. *IEEE Transactions on Cybernetics* 52(7): 6434–6441.

Xu Z, Sheng Y, He L and Wang Y (2016) Review on knowledge graph techniques. *Journal of University of Electronic Science and Technology of China* 45(4): 589–606.

y Arcas BA (2022) Do large language models understand us? *Daedalus* 151(2): 183–197.

Yang H, Xiao L, Zhu R, Liu Z and Chen J (2024) An llm supported approach to ontology and knowledge graph construction. In: *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, pp. 5240–5246.

Yin M, Tang L, Webster C, Yi X, Ying H and Wen Y (2024) A deep natural language processing-based method for ontology learning of project-specific properties from building information models. *Computer-Aided Civil and Infrastructure Engineering* 39(1): 20–45.

Ying W (2009) *Design methodology for ontology-based multi-agent applications (MOMA)*. PhD Thesis, UNSW Sydney.

Ymele CAA and Jiomekong A (2024) Tsotsalearning at llms4ol tasks a and b: Combining rules to large language model for ontology learning. In: *Open Conference Proceedings*, volume 4. pp. 65–76.

Yu YT and Hsu CC (2011) A structured ontology construction by using data clustering and pattern tree mining. In: *2011 International Conference on Machine Learning and Cybernetics*, volume 1. IEEE, pp. 45–50.

Yunianta A, Barukah OM, Yusof N, Musdholifah A, Jayadiyanti H, Dengen N, Othman MS et al. (2017) The ontology-based methodology phases to develop multi-agent system (ommas). In: *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. IEEE, pp. 1–6.

Yunianta A, Basori AH, Prabuwono AS, Bramantoro A, Syamsuddin I, Yusof N, Almagrabi AO and Alsubhi K (2019) Ontodi: The methodology for ontology development on data integration. *(IJACSA) International Journal of Advanced Computer Science and Applications* .

Zaouga W and Rabai LBA (2021) A decision support system for project risk management based on ontology learning. *International Journal of Computer Information Systems and Industrial Management Applications* 13: 11–11.

Zengeya T and Fonou-Dombeu JV (2024) A review of state of the art deep learning models for ontology construction. *IEEE Access* 12: 82354–82383.

Zhang B, Carriero VA, Schreiberhuber K, Tsaneva S, González LS, Kim J and de Berardinis J (2024) Ontochat: a framework for conversational ontology engineering using language models. In: *European Semantic Web Conference*. Springer, pp. 102–121.

Zhang L, Chen H, Zhang Y and Chen T (2018) Fine construction of hiv protein ontology. In: *2018 3rd International Conference on Information Systems Engineering (ICISE)*. IEEE, pp. 82–88.

Zhang W, Cao Y and Zhang J (2026) Knowledge graph-enhanced large language models for explainable causal analysis of crane accidents. *Reliability Engineering & System Safety* 276: 112859. DOI:https://doi.org/10.1016/j.ress.2026.112859. URL https://www.sciencedirect.com/science/article/pii/S0951832026006708.

Zhang Y, Saberi M and Chang E (2017) Semantic-based lightweight ontology learning framework: a case study of intrusion detection ontology. In: *Proceedings of the international conference on web intelligence*. pp. 1171–1177.

Zhao D, Sun J, Chen X, Bo X, Yi M and Xia L (2021) Semi-automatic construction method of power safety ontology based on ar-k-means. In: *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*. IEEE, pp. 59–64.

Zhao L, Zhang X, Wang K, Feng Z and Wang Z (2019) Learning ontology axioms over knowledge graphs via representation learning. In: *18th International Semantic Web Conference (ISWC 2019)*. Rheinisch-Westfaelische Technische Hochschule Aachen.

Zhao W, Fu Z, Fan T and Wang J (2023) Ontology construction and mapping of multi-source heterogeneous data based on hybrid neural network and autoencoder. *Neural Computing and Applications* 35(36): 25131–25141.

Zulkipli ZZ, Maskat R and Teo NHI (2022) A systematic literature review of automatic ontology construction. *Indones. J. Electr. Eng. Comput. Sci* 28(2): 878.

Appendix

In this section, we report the supplementary data to trace the proposed results. This appendix includes: the clusters for the best k obtained from the manual ontology classification in Table 16, the created semi-automatic methodologies dataset in Table 17, and the created automatic methodologies dataset in Tables 18 and 19.

Cluster	Paper(s)
0	Plinius van der Vet et al. (1994), Sensus Swartout et al. (1996), Mikrokosmos O'Hara et al. (1998), Daponte and Falquet (2018), Fawei et al. (2018)
1	PhysSys Borst et al. (1995), Jacksi (2019), Tang et al. (2023a)
2	Samod Peroni (2016), Abdelghany et al. (2019), Linked Open Terms Poveda-Villalón et al. (2022)
3	Onions Gangemi et al. (1996), $(KA)^2$ Benjamins and Fensel (1998), Hcome Kotis and Vouros (2006), Dogma Jarrar and Meersman (2009)
4	Grüninger & Fox Gruninger (1995), Diligent Pinto et al. (2004)
5	Tove Fox (1992), Uschold & King Uschold and King (1995), Methontology Fernández-López et al. (1997), Ontology Development 101 Noy et al. (2001)
6	NeOn Suárez-Figueroa et al. (2011), Grounded Ontology Nabi and Asif (2015)
7	eXtreme Hristozova and Sterling (2002), Saad and Shaharin (2016)
8	Debruyne and Meersman (2012), Bautista-Zambrana (2015)
9	Idef5 Peraketh et al. (1994)
10	Kads Wielinga et al. (1992), Scim John et al. (2018)
11	Li et al. (2009), Yamo Dutta et al. (2015)
12	Upon De Nicola et al. (2005), UponLite De Nicola and Missikoff (2016)
13	Cyc Lenat et al. (1985), Dolce Gangemi et al. (2003), Li and Alian (2018), OntoDi Yunianta et al. (2019), Oda Sharma and Jain (2024), Tailhardat et al. (2024)
14	On-To-Knowledge Sure et al. (2003), Zhang et al. (2018), Olivares-Alarcos et al. (2022)
15	DynamOnt Gahleitner et al. (2005)
16	OmMas Yunianta et al. (2017), Ahmad et al. (2018)

Table 16. Best clusters of manual ontology creation papers.

Paper	Manual	Automatic
Wang et al. Wang et al. (2017a)	Creation and experts' revision	Spanning Tree
Rani et al. Rani et al. (2017)	Creation and experts' revision	SVD
Qiu et al. Qiu et al. (2017)	Middle-out	Semantic Similarity
Fawei et al. Fawei et al. (2019)	Creation and experts' revision	Stanford CoreNLP
Chandolikar et al. Chandolikar et al. (2019)	Scraping	LSTM
Dera et al. Dera et al. (2020)	Scraping	Stanford CoreNLP
Ma et al. Ma et al. (2021)	Relational database	Model checking
Zhao et al. Zhao et al. (2021)	Scraping	Semantic Similarity
ten Haaf et al. ten Haaf et al. (2021)	Scraping	Clustering
ElAssy et al. ElAssy et al. (2022)	Creation and experts' revision	Data Mapping
Tang et al. Tang et al. (2023b)	Ontology Development 101 Noy et al. (2001)	LLM-based
Mesmia et al. Mesmia and Mouhoub (2023)	Scraping	Data Mapping
Pan et al. Pan et al. (2023)	NeOn Suárez-Figueroa et al. (2011)	Data Mapping
Sobhogol et al. Sobhgol et al. (2023)	Relational database	Data Mapping
Kommineni et al. Kommineni et al. (2024)	Creation and experts' revision	LLM-based
Doumanas et al. Doumanas et al. (2024)	HCOME Kotis and Vouros (2006)	LLM-based
Huang et al. Huang et al. (2024b)	UponLite De Nicola and Missikoff (2016)	LLM-based
Wu et al. Wu et al. (2025)	Creation and experts' revision	Association Rules
Gadusu et al. Gadusu et al. (2025)	McGinty et al. McGinty (2018)	Data Mapping
Lichtnow et al. Lichtnow et al. (2025)	Ontology Development 101 Noy et al. (2001)	LLM-based

Table 17. Classification of semi-automatic methodologies.

Paper	Tasks	Strategies
Casteleiro et al. Casteleiro et al. (2016)	Relation Extraction	Distributional Semantics
Hamano et al. Hamano et al. (2017)	Relation Extraction	Specific
Albukhitan et al. Albukhitan et al. (2017)	Concept/Taxonomy Extraction	CBOW, Skip-gram
Cahyani et al. Cahyani and Wasito (2017)	Relation Extraction, Matching	Specific
Mathews et al. Mathews and Kumar (2017)	Concept Learning, Taxonomy Extraction	Specific
Wang et al. Wang et al. (2017b)	Relation Extraction	Specific
Zhang et al. Zhang et al. (2017)	Relational Extraction	NLP Pipeline
Potoniec et al. Potoniec et al. (2017)	Taxonomy Extraction	Pattern Mining
Albarghothi et al. Albarghothi et al. (2018)	Relation Extraction	NLP Pipeline
Aggoune et al. Aggoune (2018)	Ontology Mapping	Specific
Kaushik et al. Kaushik and Chatterjee (2018)	Ontology Mapping	Specific
Petrucci et al. Petrucci et al. (2018)	Concept Learning	RNN, Encoder/Decoder
Alobaidi et al. Alobaidi et al. (2018)	Concept/Relation Extraction	NLP Pipeline
Deuna et al. De Uña et al. (2018)	Ontology Mapping	Pattern Mining, Rule-based
An et al. An and Park (2018)	Ontology Mapping	Rule-based
Zhao et al. Zhao et al. (2019)	Axiom Learning	Embedding
Capuano et al. Capuano et al. (2020)	Concept Extraction	CNN
Zaouga et al. Zaouga and Rabai (2021)	Concept Learning, Relation/Rule Extraction	NLP Pipeline, Pattern Mining
Ben et al. Ben Mahria et al. (2021)	Ontology Mapping	Specific
Chen et al. Chen and Gu (2021)	Concept/Relation Extraction	LSTM, CNN
Mhammedi et al. Mhammedi et al. (2021)	Ontology Mapping	Rule-based
Huang et al. Huang et al. (2021)	Terms clustering	Latent Dirichlet Analysis
Heidari et al. Heidari et al. (2021)	Relation Extraction	Attention-based
Heidari et al. Heidari et al. (2021)	Relation Extraction	Attention-based
Lakzaei et al. Lakzaei and Shamsfard (2021)	Ontology Mapping	Rule-based
Meriah et al. Meriah et al. (2021)	Concept Extraction	NLP Pipeline
Felin et al. Felin and Tettamanzi (2021)	Axiom Learning	Evolution, Rule-based
Xie et al. Xie et al. (2022)	Relation Extraction	Latent Dirichlet Analysis
Leshcheva et al. Leshcheva and Begler (2022)	Ontology Mapping	Rule-based
Tissaoui et al. Tissaoui et al. (2022)	Concept Extraction	Latent Dirichlet Analysis
Elnagar et al. Elnagar et al. (2022)	Relation Extraction	NLP Pipeline

Table 18. Classification of semi-automatic methodologies (2016-2022).

Paper	Tasks	Strategies
Taye et al. Taye et al. (2023)	Concept/Relation Extraction	NLP Pipeline, Transformer-based
Al-Aswadi et al. Al-Aswadi et al. (2023)	Concept Extraction	LSTM
He et al. He et al. (2023)	Ontology Alignment	Zero-shot LLM
Babaei et al. Babaei Giglou et al. (2023)	Concept/Taxonomy/Relation Extraction	Zero-shot LLM, Fine-tuned LLM
Schaeffer et al. Schaeffer et al. (2023)	Text-to-Ontology	Specific
Hertling et al. Hertling and Paulheim (2023)	Ontology Mapping	Few-shot LLM
Zhao et al. Zhao et al. (2023)	Ontology Mapping	LSTM, CNN, Encoder/Decoder
Mateiu et al. Mateiu and Groza (2023)	Text-to-Ontology	Zero-shot LLM
Hari et al. Hari and Kumar (2023)	Relation Extraction	Transformer-based
Yin et al. Yin et al. (2024)	Domain Ontology	Transformer-based
Dong et al. Dong et al. (2024)	Concept/Relation Extraction	Zero-shot LLM, Fine-tuned LLM
Yang et al. Yang et al. (2024)	Domain Ontology	Zero-shot LLM
Al-Subhi et al. Al Subhi et al. (2024)	Instantiation	Transformer-based
Basheer et al. Basheer and Alkhatib (2024)	Relation Extraction	Skip-gram
Mai et al. Mai et al. (2024)	Concept/Taxonomy/Relation Extraction	Few-shot LLM
Toro et al. Toro et al. (2024)	Concept Discovery, Relation Extraction	RAG, Zero-shot LLM
Lo et al. Lo et al. (2024)	Framework	Fine-tuned LLM
Pisu et al. Pisu et al. (2024)	Relation Extraction	Transformer-based
Abolhasani et al. Abolhasani and Pan (2024)	Text-to-Ontology	Zero-shot LLM
Coutinho et al. Coutinho (2024)	Framework	Few-shot LLM
Bischof et al. Bischof et al. (2024)	Annotations generation	Zero-shot LLM
Fathallah et al. Fathallah et al. (2024b)	Domain Ontology	Zero-shot LLM
Fathallah et al. Fathallah et al. (2024a)	Framework	Zero-shot LLM
Fathallah et al. Fathallah et al. (2024a)	Framework	Zero-shot LLM
Dasilva et al. da Silva et al. (2024)	Framework	Zero-shot LLM
Zhang et al. Zhang et al. (2024)	Framework	Zero-shot LLM
Huang et al. Huang et al. (2024a)	Domain Ontology	Embedding
Bakker et al. Bakker et al. (2024)	Text-to-Ontology	Zero-shot LLM
Feng et al. Feng et al. (2024)	CQ Generation, Relation Extraction, Matching	Zero-shot LLM
Sanaei et al. Sanaei et al. (2024)	Concept/Taxonomy/Relation Extraction	RAG
Goyal et al. Goyal et al. (2024)	Concept/Taxonomy/Relation Extraction	Zero/Few-shot LLM
Joachimiak et al. Joachimiak et al. (2024)	Domain Ontology	Zero-shot LLM
Ymele et al. Ymele and Jiomekong (2024)	Concept/Taxonomy/Relation Extraction	Fine-tuned LLM
Lippolis et al. Lippolis et al. (2025)	Framework	Few-shot LLM
Bosso et al. Bosso et al. (2025)	Text-to-Ontology	Few-shot LLM

Table 19. Classification of semi-automatic methodologies (2023-2025).