Overview of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026

Federica Vezzani[1], Giorgio Maria Di Nunzio[2,*], Vanessa Bonato[1] and Gianmaria Silvello[2]

[1]*Department of Linguistic and Literary Studies, University of Padua, Italy*
[2]*Department of Information Engineering, University of Padua, Italy*

Abstract

This paper presents an overview of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, held in conjunction with Multilingual Digital Terminology Today (MDTT) 2026. DETECH 2026 focuses on automatic term extraction and definition generation in the biomedical domain, with particular attention to the gut–brain interplay and to two related subdomains: mental health and Parkinson's disease. The challenge is organized around two tasks: Task A, devoted to the extraction of single-word and multi-word biomedical terms from English PubMed abstracts, and Task B, devoted to concept assignment and natural language definition generation. This overview introduces the scientific motivation of the challenge, describes its task setting and evaluation framework, and summarizes the accepted contributions. The papers presented at DETECH 2026 cover a broad methodological spectrum, including symbolic term extraction, BERT-based sequence labelling, biomedical transformer models, customized generative AI, retrieval-augmented generation, ontology-enhanced prompting, and theory-driven approaches to terminological meaning. Taken together, the contributions show that term extraction and definition generation should be treated as connected knowledge-oriented tasks requiring terminological adequacy, biomedical grounding, explainability, and expert-informed evaluation.

Keywords

Automatic term extraction, Definition generation, Biomedical terminology

1. Introduction

Specialized knowledge is increasingly accessed, processed, and reused through computational systems that operate on scientific and technical texts. In biomedical domains, this tendency is particularly visible: large collections of publications, clinical reports, terminological resources, ontologies, and knowledge bases are routinely used to support information retrieval, text mining, evidence synthesis, decision support, and expert communication. However, the effectiveness of these systems depends not only on the availability of textual data, but also on the capacity to identify the domain-specific concepts expressed in such data and to make their meaning explicit in a reliable and interpretable way. Terminology therefore remains a central component of biomedical language technologies, because terms provide linguistic access to specialized concepts, while definitions support conceptual clarification, knowledge organization, and communication across expert and non-expert communities.

Automatic term extraction (ATE) has long addressed the problem of identifying lexical units that designate domain-relevant concepts in specialized corpora. In biomedical texts, this task is especially demanding because terms may occur as single-word units, complex multi-word expressions, abbreviations, variants, coordinated structures, or nested expressions. Moreover, the distinction between general language and specialized usage is often context-dependent. A lexical unit may acquire terminological status only within a specific domain, corpus, or communicative setting. This makes term extraction more than a surface-level recognition problem: it requires the modelling of termhood, unithood, domain relevance, and textual context [1]. Traditional linguistic and statistical approaches have progressively been complemented by supervised sequence labelling, transformer-based models, biomedical language models, and

International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, June 24, 2026, Zadar, Croatia
*Corresponding author.
federica.vezzani@unipd.it (F. Vezzani); giorgiomaria.dinunzio@unipd.it (G. Di Nunzio); vanessa.bonato@phd.unipd.it (V. Bonato); gianmaria.silvello@unipd.it (G. Silvello)
0000-0003-2240-6127 (F. Vezzani); 0000-0001-9709-6392 (G. Di Nunzio); 0009-0002-9918-282X (V. Bonato); 0000-0003-4970-4554 (G. Silvello)

more recently large language models (LLMs), but the problem of robust and explainable terminology extraction remains open, especially in highly specialized and interdisciplinary domains [2, 3].

Definition extraction and definition generation represent a complementary challenge. If term extraction identifies the linguistic units that provide access to specialized concepts, definition-oriented tasks aim to make explicit the conceptual content associated with those units. Terminological definitions are not merely paraphrases of words; they are expected to delimit concepts within a domain-specific conceptual system, often by identifying a superordinate concept and the characteristics that distinguish the defined concept from related ones. In computational settings, definitions are relevant for terminology management, ontology learning, concept normalization, knowledge base construction, retrieval-augmented generation, and explainable artificial intelligence. Yet their automatic production is difficult to evaluate, because fluency alone is insufficient: a generated definition must also be conceptually correct, terminologically precise, domain-appropriate, and useful for the intended users. A closely related evaluation setting was proposed in the CLEF 2024 SimpleText Task 2, which combined difficult concept spotting, definition generation, and definition ranking for scientific texts [4, 5].

The International Workshop on Definition and Term Extraction Challenge (DETECH) 2026 was organized to address these issues through a shared task and workshop setting. DETECH 2026 is a HEREDITARY event and is part of the broader activities of the EU-supported project "HetERogeneous sEmantic Data integratIon for the guT-bRain interplaY",[1] which promotes research on semantic integration and biomedical text processing for the gut–brain interplay. The challenge focuses on biomedical terminology in the domain of the gut–brain interplay, an interdisciplinary area that connects gastroenterology, neuroscience, microbiology, genetics, mental health, and neurodegenerative disease research. This domain is particularly suitable for a terminology-oriented challenge because it involves rapidly evolving scientific knowledge, heterogeneous conceptual relations, and complex multi-word expressions that cross traditional disciplinary boundaries. The workshop is organized around two related tasks: Task A, devoted to automatic term extraction from English biomedical abstracts, and Task B, devoted to concept assignment and definition generation. Together, the two tasks reflect a broader view of computational terminology, in which identifying terms and defining the concepts they designate are treated as connected stages of terminology processing.

The scientific motivation of DETECH 2026 is twofold. First, the workshop aims to provide a benchmark for evaluating systems that extract biomedical terms from specialized corpora. This involves assessing not only whether systems can recover relevant term occurrences and term types, but also how they handle the boundary and granularity problems that characterize biomedical terminology. Second, the workshop aims to stimulate research on automatic definition generation for specialized concepts. This requires moving beyond generic text generation toward methods that can exploit contextual evidence, biomedical knowledge sources, terminological constraints, and evaluation procedures sensitive to conceptual adequacy. In this sense, DETECH 2026 contributes to the broader development of terminology-aware natural language processing and to current discussions on the role of structured terminology in explainable and knowledge-grounded artificial intelligence.

The accepted papers of DETECH 2026 reflect the methodological diversity of current research in automatic term extraction and definition generation. Some contributions focus primarily on Task A, exploring customized generative AI systems, BERT-enhanced terminology extraction tools, and comparisons between symbolic and neural approaches. Other contributions address both tasks through integrated pipelines that combine biomedical transformer models, retrieval mechanisms, structured prompting, ontology-enhanced generation, or theory-driven contextual representations. A further contribution provides a structured review of terminological definitions, situating definition generation within the longer tradition of terminology theory, specialized lexicography, knowledge organization, and computational terminology. Taken together, these papers show that the field is moving toward hybrid approaches in which linguistic knowledge, domain resources, neural modelling, and generative methods are combined rather than treated as mutually exclusive alternatives.

This overview paper introduces the scientific context of DETECH 2026 and presents the main themes

[1] https://hereditary-project.eu/

emerging from the accepted contributions. Section 2 describes the DETECH 2026 challenge, its tasks, domain, data, and evaluation setting. Section 2.1 and 2.2 summarizes the accepted papers and highlights their methodological contributions. Section 3 discusses trends and open challenges emerging from the workshop, including term-boundary detection, concept normalization, definition quality, grounding, explainability, and evaluation.

2. The DETECH 2026 Challenge

The International Workshop on Definition and Term Extraction Challenge (DETECH) 2026 was designed as a shared evaluation setting for computational terminology in the biomedical domain. Its main objective is to bring together research on automatic term extraction and definition generation, two tasks that are often studied separately but are closely connected in terminological practice. Term extraction identifies the lexical units that provide access to specialized concepts, whereas definition generation aims to make explicit the conceptual content associated with those units. By combining these tasks within the same challenge, DETECH 2026 promotes an integrated view of terminology processing, in which the recognition of terms and the production of definitions are treated as complementary stages in the construction, validation, and reuse of domain knowledge.

The 2026 edition focuses on the gut–brain interplay, an interdisciplinary biomedical domain at the intersection of gastroenterology, neuroscience, microbiology, genetics, mental health, and neurodegenerative disease research. This domain provides a particularly suitable testbed for terminology-oriented natural language processing because it involves complex conceptual relations across several areas of biomedical knowledge. Research on the gut–brain axis addresses interactions between gut microbiota, gastrointestinal processes, neurological conditions, psychiatric disorders, inflammation, immune mechanisms, metabolic factors, and clinical outcomes. As a result, the terminology of the domain includes general biomedical terms, highly specialized expressions, multi-word units, abbreviations, and emerging conceptual combinations that are not always stabilized in terminological resources.

DETECH 2026 is organized around two subtasks:

- Task A, Term Extraction, requires systems to identify relevant single-word and multi-word terms from English biomedical texts concerning the gut–brain interplay. The task is formulated at both the occurrence level and the type level. At the occurrence level, systems are expected to identify the mentions of terms as they appear in the documents. At the type level, systems are expected to recover the distinct terminology of the domain after light normalization of surface forms. This dual perspective reflects a central issue in automatic terminology extraction: a system must be able to recognize terms in context, but it must also provide a useful inventory of domain-relevant lexical units.
- Task B, Definition Generation, requires systems to produce natural language definitions for the concepts designated by the extracted terms. The task moves from term recognition to conceptual description and therefore introduces a different set of requirements. A generated definition should not only be linguistically fluent, but also terminologically adequate, conceptually informative, and consistent with the biomedical domain. In line with terminological theory and definition-writing standards, the expected output is oriented toward intensional definitions, where a concept is described by locating it within a broader conceptual category and by specifying relevant distinguishing characteristics. Systems may rely on corpus-based evidence, biomedical resources, retrieval mechanisms, prompting strategies, neural sequence-to-sequence models, large language models, or hybrid approaches.

The challenge data are based on PubMed[2] abstracts and are organized around two thematic domains: mental health and Parkinson's disease. These domains were selected because they represent two important areas in which the gut–brain interplay has become increasingly relevant. Mental health research involves terms related to psychiatric disorders, gut microbiota, behavioural conditions, inflammation,

[2]https://pubmed.ncbi.nlm.nih.gov/

metabolism, and psychopharmacology. Parkinson's disease research involves terms related to neurodegeneration, motor and non-motor symptoms, gastrointestinal dysfunction, microbiome composition, alpha-synuclein, inflammation, and disease mechanisms. The use of two related but distinct subdomains makes it possible to evaluate whether systems can generalize across partially overlapping biomedical contexts while still capturing the specific terminology of each domain.

The challenge material was distributed through the DETECH 2026 GitHub repository.[3] The repository provides the datasets, the annotation protocol, and the task-specific folders for Task A and Task B. For Task A, it includes document lists, PubMed abstract files, and term annotation files for the two domains. For Task B, it includes the abstract subsets, mention-to-concept supervision files, and gold intensional definitions used for training. The repository also documents the expected submission setting, including the distinction between term occurrences and term types for Task A, and the link between mentions, normalized concepts, and intensional definitions for Task B.

The evaluation of Task A is based on Micro-F1 and Type-F1. Micro-F1 measures how consistently a system identifies term occurrences across the corpus and is therefore sensitive to mention-level extraction and exact boundary detection. Type-F1 evaluates the ability of a system to identify distinct term types and is therefore closer to the construction of a terminology inventory. The use of both measures is important because occurrence-level and type-level performance may diverge. A system may identify repeated occurrences of frequent terms while missing rare but relevant terminology, or conversely recover many distinct terms while failing to detect all their contextual mentions. The two scores therefore provide complementary views of extraction performance.

The evaluation of Task B combines automatic metrics and qualitative assessment. BLEU and BERTScore are used to compare generated definitions with reference definitions from complementary perspectives: lexical overlap and semantic similarity. However, definition generation in specialized domains cannot be fully assessed through automatic metrics alone. A definition may achieve a relatively high similarity score while remaining conceptually incomplete, terminologically imprecise, or insufficiently informative for domain use. Conversely, a valid definition may differ lexically from a reference formulation while preserving the relevant conceptual content. For this reason, DETECH 2026 also includes additional manual or qualitative checks aimed at evaluating the informativeness and linguistic quality of generated definitions.

Participants were allowed to submit up to five runs per subtask. The challenge accepted a broad range of methodological approaches, including linguistic processing, statistical methods, supervised neural models, transformer-based sequence labelling, external lexicons, biomedical ontologies, retrieval-based methods, and generative AI systems. External resources such as pre-trained models, terminology databases, lexicons, and ontologies were permitted, provided that their use was documented. Manual runs were accepted for exploratory comparison but were not considered for ranking. This design was intended to encourage methodological diversity while preserving the transparency required for a meaningful shared-task evaluation.

2.1. Main Results

A total of five teams participated in the DETECH 2026 shared task, submitting 42 runs across the two evaluation tasks. Task A, dedicated to Automatic Term Extraction, received 30 runs from five participating teams, while Task B, focusing on Concept Identification and Definition Generation, received 12 runs from two teams. All submitted systems were evaluated separately on the Mental Health and Parkinson subcollections. In the following, we present an overview of the results obtained on each task, discussing the performance of participating systems and the main trends emerging from the evaluation.

Figure 1 shows the distribution of Micro-F1 and Type-F1 scores across all submitted runs for Task A. The results reveal moderate variability among participating systems, with median scores generally ranging between 0.45 and 0.60. The Mental Health subcollection exhibits slightly higher median performance than the Parkinson subcollection for both Micro-F1 and Type-F1, suggesting that terminology

[3]https://github.com/gmdn/DETECH2026/

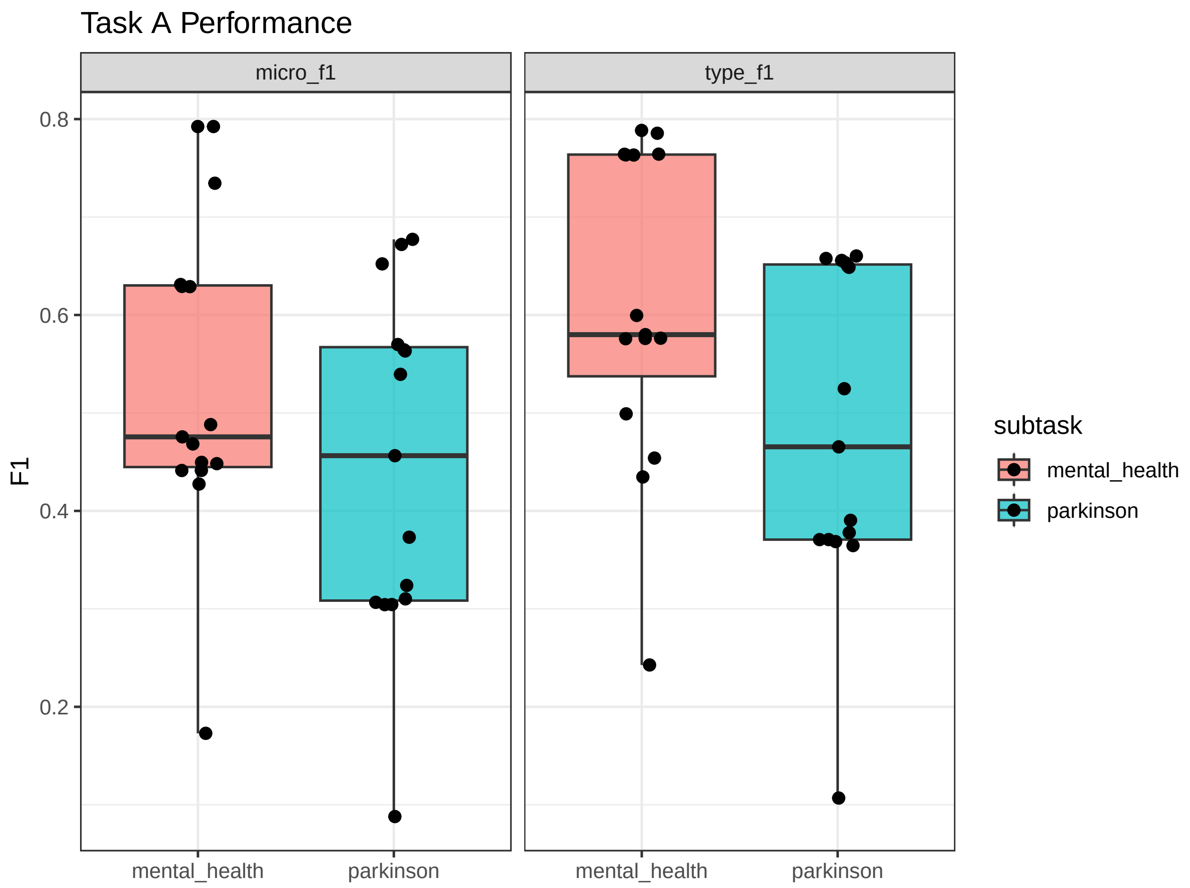

Figure 1: Distribution of Micro-F1 and Type-F1 scores across all submitted runs for Task A. Results are shown separately for the Mental Health and Parkinson subcollections. The boxplots summarize the distribution of system performance, while individual points correspond to submitted runs.

extraction may be comparatively easier in that domain. At the same time, the spread of scores indicates that methodological choices substantially affect performance, with the best systems achieving Type-F1 values close to 0.8. These results confirm both the difficulty of the task and the potential of current approaches to identify specialized terminology in biomedical literature.

Figure 2 presents the distribution of the main Task B evaluation measures across all submitted runs. Overall, the results indicate that systems were generally able to generate semantically appropriate definitions for correctly identified concepts, as reflected by the consistently high BERTScore values, which remain above 0.92 for both subcollections. In contrast, BLEU-1 and ROUGE-L scores are substantially lower, highlighting the lexical variability of generated definitions and the frequent use of paraphrasing strategies. Interestingly, the Parkinson subcollection achieves slightly higher semantic similarity scores, whereas the Mental Health subcollection exhibits marginally higher lexical overlap with the reference definitions. The relatively limited variability observed for BERTScore suggests that most systems converged towards comparable semantic quality, while the wider spread of BLEU and ROUGE scores reflects differences in the wording and formulation strategies adopted by participants. These findings confirm the importance of combining lexical and semantic evaluation measures when assessing automatically generated biomedical definitions.

2.2. Overview of the Approaches by Participant

The papers accepted at DETECH 2026 reflect the methodological diversity of current research in automatic term extraction and definition generation. They range from symbolic and linguistically motivated systems to transformer-based sequence labelling, customized generative AI, retrieval-augmented gener-

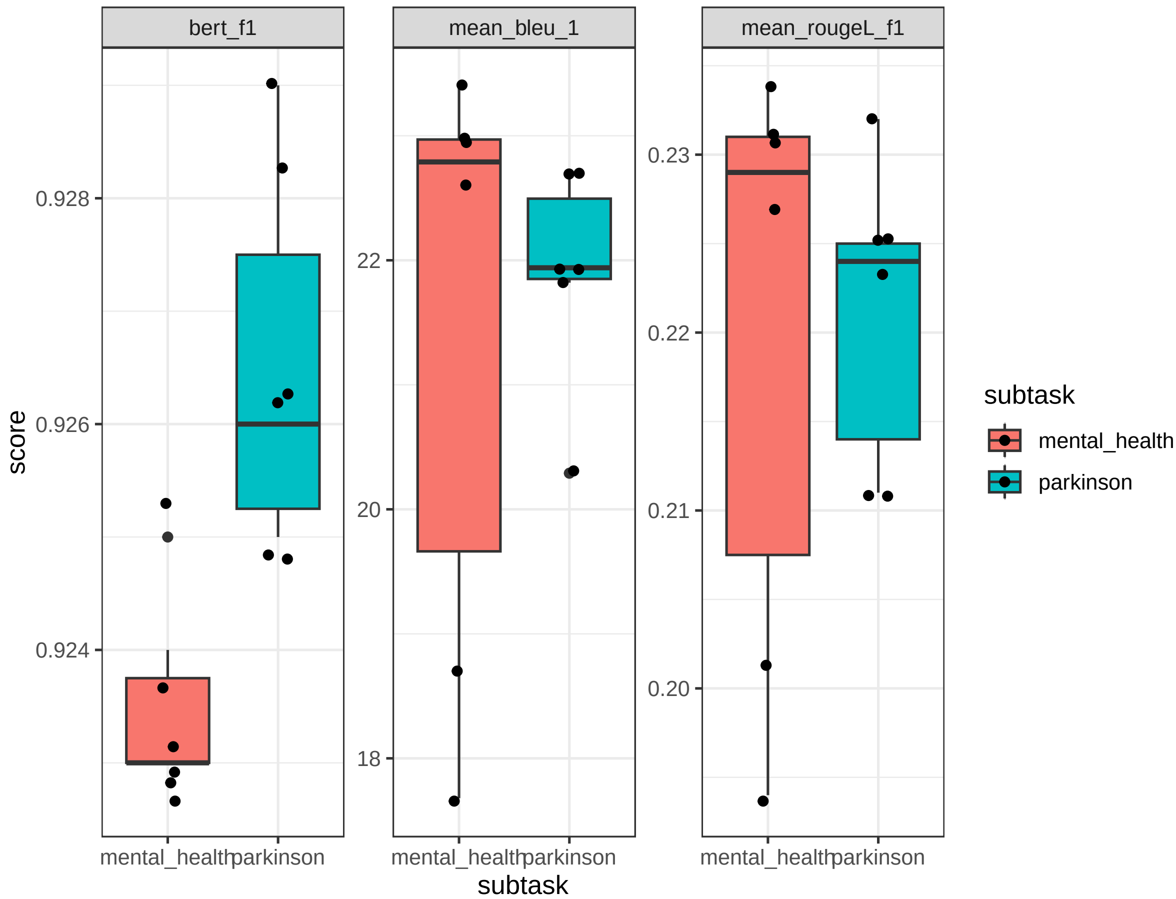

Figure 2: Distribution of Task B definition-generation scores across submitted runs. Results are reported separately for the Mental Health and Parkinson subcollections using BLEU-1, ROUGE-L F1, and BERTScore F1. Individual points correspond to submitted runs, while boxplots summarize the score distributions.

ation, and theoretically informed approaches to terminological meaning. This diversity is particularly relevant for a challenge devoted to biomedical terminology, since the extraction of terms and the generation of definitions require both computational effectiveness and terminological adequacy. The accepted contributions also show that the two DETECH tasks can be addressed either separately or as part of an integrated terminology-processing pipeline.

A first group of papers focuses primarily on Task A, namely the extraction of biomedical terms from the mental health and Parkinson's disease corpora. Ortiz-Garduño and Castillo-Pérez [6] present *GutBrainTerm_Extractor*, a generative-AI-based system developed by the UGR-TermiGenAI team. The system was implemented using ChatGPT-5.4 and configured through customized internal instructions aligned with the DETECH annotation protocol. The authors processed the dataset in five separate runs, distinguishing between the mental health and Parkinson's disease subsets, in order to reduce the risk of omissions and hallucinated outputs associated with very large inputs. Their results show a precision-oriented profile, with stronger performance on the mental health corpus than on the Parkinson's disease corpus. The paper is significant because it explores the use of a general-purpose generative AI system as an accessible tool for terminology extraction, while also emphasizing the need for task-specific constraints, careful prompt design, and control of non-deterministic behaviour.

López-Sánchez, Morales-Hurtado, Vázquez, Morales-Moreno, and Rodríguez Vázquez [7] present a system that adapts TBXTools to the DETECH Task A setting by integrating a BERT-based approach. TBXTools is originally a terminology extraction tool based on statistical and linguistic methods, and the paper extends it by modelling automatic terminology extraction as a token-level sequence labelling task using BIO tagging. The authors compile and exploit reference term lists from medical and health-related sources, including resources related to genomics, health, mental health, Parkinson's disease, gastrointestinal terminology, and biomedical databases. In this contribution, BERT is used as a filtering mechanism applied to terminology extraction, with the aim of improving precision and recall in comparison with more traditional extraction settings. The paper is especially relevant for DETECH because it connects established terminology-extraction software with contemporary transformer-based modelling, showing a possible path for updating terminology tools without abandoning their terminological orientation.

Hamon [8] describes the LISN system for Task A, which compares a rule-based term extraction approach with neural sequence-tagging approaches. The symbolic component relies on YATEA, a state-of-the-art term extractor, whereas the neural component treats term extraction as a sequential tagging problem. Two different representations are explored in the machine-learning setting, including an approach based on tagging the term contexts rather than only the term components. The results reported by the author indicate that the symbolic approach achieved better performance than the machine-learning approach in the DETECH setting, although context-oriented tagging produced promising results. This contribution is important because it provides a useful counterpoint to purely neural approaches: in specialized domains, well-established linguistic and symbolic systems may remain competitive, especially when the target phenomenon involves complex multi-word terms, syntactic regularities, and domain-specific candidate filtering.

A second group of papers addresses both DETECH tasks through integrated pipelines that connect term extraction, concept assignment, and definition generation. Hosseini-Kivanani and Resi [9] present *TermHunter*, a cascaded system combining neural biomedical term extraction and ontology-enhanced definition generation. For Task A, the system combines BiomedBERT-large and BioLinkBERT-large sequence taggers through a confidence-calibrated token ensemble, and compares the neural baseline with several linguistic post-processing modules for boundary correction, confidence filtering, lexical filtering, abbreviation recovery, and related refinements. For Task B, the strongest Task A extractor is used to identify concepts in the definition-generation abstracts, after which definitions are produced with BioBART-v2-large, concept-type-aware prompting, retrieval augmentation, self-consistency decoding, and optional LLM-based re-ranking. The paper shows that controlled neural extraction and compact retrieval-grounded generation can be more effective than extensive post-hoc correction in this shared-task setting. It also illustrates the growing importance of modular pipelines in which extraction, retrieval, generation, and evaluation are explicitly connected.

Burgos, Tamayo Herrera, Díaz, and Pérez-Pérez [10] propose a system grounded in the Quantum Theory of Terms (QTT), applying it to both automatic term extraction and definition generation. Their approach operationalizes QTT through contextual semantic-state induction and a collapse-based coherence signal. For Task A, the system combines a fine-tuned PubMedBERT span extractor with a contextual refinement module that computes mention embeddings, induces latent states with clustering, and estimates collapse from centroid distance. For Task B, extracted mentions are mapped to concepts using training-derived normalization resources, and definitions are generated with a large language model using collapse-guided contextual evidence. This paper is distinctive because it attempts to connect a theory of terminological meaning with computational modelling. Rather than treating terms as stable lexical objects only, the authors model them as contextual semantic entities whose meaning is stabilized through use. In the DETECH setting, this provides an original perspective on the relation between term occurrence, concept assignment, contextual evidence, and definition generation.

A final accepted contribution provides a broader conceptual and historical framework for the definition-generation side of the challenge. Di Nunzio [11] presents a structured review of research on terminological definitions, tracing the transition from theoretical and methodological approaches to computational and generative perspectives. The paper examines the role of definitions in terminology theory, specialized lexicography, knowledge organization, ontology learning, automatic definition extraction, definition generation, and terminology-augmented generation. Its main relevance for DETECH lies in the clarification that terminological definitions cannot be evaluated only as fluent textual outputs. They are expected to delimit specialized concepts, support conceptual organization, and provide domain-appropriate knowledge. This review therefore complements the system papers by situating Task B within a longer research tradition concerned with definition quality, intensional structure, conceptual adequacy, and expert validation.

3. Trends and Open Challenges

The contributions to DETECH 2026 confirm that automatic term extraction and definition generation are increasingly shaped by the convergence of computational terminology, biomedical natural language processing, and generative artificial intelligence. At the same time, they show that the availability of powerful language models does not eliminate the classical problems of terminology work. On the contrary, issues such as termhood, unithood, conceptual delimitation, definition quality, and expert validation become even more important when extraction and generation are delegated, even partially, to automatic systems.

A first trend concerns the coexistence of symbolic, statistical, neural, and generative approaches. The accepted papers do not point toward the replacement of earlier terminology extraction methods by a single dominant paradigm. Instead, they suggest that different methodological families capture different aspects of the problem. Rule-based and linguistically informed systems remain valuable because biomedical terms often exhibit recurrent morphosyntactic patterns, especially in complex noun phrases. Statistical and corpus-based evidence remains useful for ranking and filtering candidate terms. Transformer-based sequence labelling improves contextual recognition and can model term boundaries directly from annotated examples. Generative AI and LLM-based systems add flexibility, especially when task-specific instructions, prompting strategies, or retrieval mechanisms are used. The open challenge is therefore not simply to choose between symbolic and neural methods, but to design hybrid systems in which linguistic constraints, corpus evidence, biomedical knowledge, and contextual representations support each other.

A second challenge concerns term boundaries and granularity. Biomedical terminology includes long multi-word expressions, nested terms, coordinated structures, abbreviations, acronyms, variants, and partially overlapping expressions. These phenomena are difficult because they affect both occurrence-level and type-level evaluation. At the occurrence level, a system may identify the correct lexical material but fail because the predicted boundary is too broad or too narrow. At the type level, small differences in normalization may lead to different term inventories. This problem is especially visible in interdisciplinary domains such as the gut–brain interplay, where a valid term may combine elements from microbiology, neuroscience, psychiatry, genetics, and gastroenterology. Future work should therefore pay more attention to annotation criteria, boundary typologies, nested term representation, and evaluation measures capable of distinguishing severe conceptual errors from minor boundary variations.

A third trend concerns the increasing role of domain-specific biomedical language models and external knowledge resources. Several systems rely on biomedical transformer encoders, curated lexicons, ontologies, thesauri, terminology databases, or retrieval mechanisms. This confirms that general-purpose language modelling is often insufficient for specialized terminology processing. Biomedical term extraction and definition generation require sensitivity to domain-specific lexical patterns, conceptual relations, abbreviations, and normalized forms. However, the use of external resources also introduces problems of coverage, provenance, compatibility, and bias. Resources may be incomplete, may reflect different conceptual granularities, or may not be fully aligned with the target corpus and annotation protocol. A major open challenge is therefore to integrate external biomedical knowledge in a transparent and reproducible way, making explicit which resources are used, how they influence system output, and how conflicts between corpus evidence and external knowledge are resolved.

A fourth issue concerns the connection between term extraction, concept assignment, and definition generation. DETECH 2026 treats these operations as related tasks, and the accepted papers show that errors propagate across them. If a system extracts an incorrect term, selects an overly broad or overly narrow span, or fails to normalize variants correctly, the generated definition may be affected even when the generation model is fluent. Conversely, a system that generates plausible definitions may still fail to provide terminologically valid knowledge if the underlying concept assignment is unstable. This suggests that future systems should make the intermediate stages of the pipeline more explicit. Term occurrence, term type, normalized concept, contextual evidence, and generated definition should be represented as connected but distinguishable layers, so that errors can be diagnosed and corrected at the appropriate level.

Definition generation raises specific challenges that are not reducible to text generation quality. In specialized domains, definitions must satisfy requirements of conceptual adequacy, terminological precision, domain relevance, and communicative usefulness. Automatic metrics such as BLEU and BERTScore provide useful approximations, but they cannot fully determine whether a definition correctly delimits a biomedical concept. A generated definition may be fluent and semantically similar to a reference while omitting a crucial distinguishing characteristic, introducing a misleading relation, or confusing neighbouring concepts. Conversely, an acceptable terminological definition may differ substantially from the reference wording. This makes expert assessment and qualitative evaluation indispensable. Future shared tasks should continue to combine automatic metrics with human evaluation, possibly by developing rubrics that assess genus identification, distinguishing characteristics, domain adequacy, non-circularity, factual correctness, and linguistic clarity.

A fifth challenge concerns explainability and trust. Biomedical terminology is often used in contexts where errors can have practical consequences for information access, evidence synthesis, translation, education, and expert communication. Systems that extract terms or generate definitions should therefore provide some account of why a term was selected, how a concept was normalized, and which evidence supported a definition. Retrieval-augmented generation, ontology-enhanced prompting, confidence scores, linguistic post-processing, and contextual evidence selection are possible strategies for improving transparency. However, explainability should not be treated only as an interface feature added after prediction. It should be built into the modelling pipeline through explicit links between textual evidence, candidate terms, conceptual resources, and generated outputs.

A sixth issue is reproducibility. The DETECH 2026 papers show increasing interest in LLM-based and GenAI-based approaches, but such systems can be difficult to reproduce because of model updates, non-deterministic behaviour, proprietary interfaces, hidden training data, and prompt sensitivity. This is especially problematic in shared-task settings, where systems are expected to be comparable and where the same method should ideally produce stable results. Future work should document prompts, internal instructions, model versions, temperature settings, retrieval resources, post-processing rules, and any manual interventions. When proprietary LLMs are used, authors should explicitly discuss the limits of reproducibility and the measures adopted to control variability.

Finally, DETECH 2026 points to the need for a closer dialogue between terminology theory and computational practice. Many technical systems implicitly make assumptions about what a term is, how a concept should be delimited, and what counts as a good definition. These assumptions are not merely implementation details: they affect annotation, modelling, evaluation, and interpretation of results. Terminology studies provide concepts such as termhood, unithood, concept orientation, intensional definition, definitional adequacy, and domain specificity, all of which remain relevant for computational systems. Conversely, computational experiments expose practical difficulties that can refine theoretical models, especially in relation to variation, granularity, contextual meaning, and the operationalization of definitional quality. The future of the field therefore depends on sustained interaction between terminologists, biomedical experts, NLP researchers, and information scientists.

Overall, the main lesson emerging from DETECH 2026 is that automatic term extraction and definition generation should be treated as knowledge-oriented tasks rather than as isolated text-processing problems. Their success depends on the ability to connect textual evidence, terminological theory, biomedical domain knowledge, computational modelling, and evaluation. This connection is particularly important in interdisciplinary domains such as the gut–brain interplay, where terminology is still evolving and where the same expression may carry different levels of specificity depending on context. DETECH 2026 therefore identifies a research agenda in which extraction, normalization, definition generation, grounding, and expert validation are developed as mutually dependent components of terminology-aware biomedical NLP.

This work has been partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074, and it is part of the initiatives of the Center for Studies in Computational Terminology (CENTRICO) of the University of Padua and in the research directions of the Italian Common Language Resources and Technology Infrastructure CLARIN-IT. This work has also been partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme through the project *Migrating Commercial Law and Language. Rethinking Lex Mercatoria (11th–17th cent.)* (MICOLL), Grant Agreement No. 101002084. This work has also been partially supported by the Croatian Science Foundation (Grant No. IP-2025-02-3140 GenAI-LT).

Acknowledgments

This work has been partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074, and it is part of the initiatives of the Center for Studies in Computational Terminology (CENTRICO) of the University of Padua and in the research directions of the Italian Common Language Resources and Technology Infrastructure CLARIN-IT.

Declaration on Generative AI

During the preparation of this work, the author(s) used Chat-GPT-5 in order to: Grammar and spelling check. After using these tools, the author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] K. Kageura, B. Umino, Methods of automatic term recognition: A review, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 3 (1996) 259–289. doi:10.1075/term.3.2.03kag.

[2] G. Di Nunzio, S. Marchesin, G. Silvello, A systematic review of Automatic Term Extraction: What happened in 2022?, Digital Scholarship in the Humanities 38 (2023) i41–i47. doi:10.1093/llc/fqad030.

[3] H. T. H. Tran, M. Martinc, J. Caporusso, J. Delaunay, A. Doucet, S. Pollak, Recent Advances in Automatic Term Extraction: A Comprehensive Survey, ACM Comput. Surv. 58 (2026) 226:1–226:35. doi:10.1145/3787584.

[4] G. M. Di Nunzio, F. Vezzani, V. Bonato, H. Azarbonyad, J. Kamps, L. Ermakova, Overview of the CLEF 2024 SimpleText Task 2: Identify and Explain Difficult Concepts, in: G. Faggioli, N. Ferro, P. Galuščáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), volume 3740 of *CEUR Workshop Proceedings*, CEUR, Grenoble, France, 2024, pp. 3129–3146.

[5] L. Ermakova, E. SanJuan, S. Huet, H. Azarbonyad, G. M. Di Nunzio, F. Vezzani, J. D'Souza, J. Kamps, Overview of the CLEF 2024 SimpleText Track, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, G. M. Di Nunzio, L. Soulier, P. Galuščáková, A. García Seco de Herrera, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer Nature Switzerland, Cham, 2024, pp. 283–307. doi:10.1007/978-3-031-71908-0_13.

[6] H. Ortiz-Gardu no, E. Castillo-P'erez, GutBrainTerm_Extractor: Generative ai for medical terminology extraction in the gut–brain domain, in: Proceedings of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.

[7] G. López-Sánchez, P. Morales-Hurtado, M. Vázquez, A. Morales-Moreno, S. Rodríguez Vázquez, Adapting TBXTools to automatic terminology extraction with BERT, in: Proceedings of the

International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.
[8] T. Hamon, Description of the LISN system for extracting terms, in: Proceedings of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.
[9] N. Hosseini-Kivanani, R. Resi, TermHunter: Neural biomedical term extraction and ontology-enhanced definition generation with structured prompting, in: Proceedings of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.
[10] D. A. Burgos, A. Tamayo Herrera, G. Díaz, C. M. Pérez-Pérez, A QTT-informed system for biomedical term extraction and definition generation, in: Proceedings of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.
[11] G. M. Di Nunzio, Defining concepts, generating knowledge: A structured review of terminological definitions, in: Proceedings of the International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.