An Ontology-Driven Knowledge Extraction Tool for Pathology Record Classification

Laura Menotti[1,*], Stefano Marchesin[1] and Gianmaria Silvello[1]

[1]Departement of Information Engineering, University of Padua, Padua, Italy

Abstract

The information in pathology diagnostic reports is often encoded in natural language. Extracting such knowledge can be instrumental in developing clinical decision support systems. However, the digital pathology domain lacks knowledge extraction systems suited to the task. One of the few examples is the Semantic Knowledge Extractor Tool (SKET), a hybrid knowledge extraction system combining a rule-based expert system with pre-trained ML models. SKET has been designed to extract knowledge from colon, cervix, and lung cancer diagnostic reports. To do so, the system employs an ontology-driven approach, where the extracted entities are linked with concepts modeled through a reference ontology, namely, the ExaMode ontology. In this work, we adapt SKET to a newer version of the ExaMode ontology and extend the method to account for an additional use case: Celiac disease. Our experimental results show that: 1) the new version of SKET outperforms the previous one on colon, cervix, and lung cancer use cases; and 2) SKET is effective on Celiac disease, confirming the ability of the system architecture to adapt to new, unseen scenarios.

Keywords

Digital Pathology, Knowledge Extraction, Expert Systems, Machine Learning

1. Introduction

Pathology revolves around studying the causes and effects of disease through the microscopic examination of tissue and human cell samples placed onto glass slides [1]. In recent years, the use of Whole Slide Images (WSIs) – obtained from the digital scanning of standard glass slides – to speed up the diagnostic process on cancer and other diseases has grown significantly [2]. Nevertheless, analyzing slides remains a time-consuming task [3]. Because of this, the use of Deep Learning (DL) models to automatically classify WSIs has risen in popularity. However, DL methods are data-hungry and require large-scale annotated datasets to be effective, which are scarce and expensive resources in the pathology domain. To overcome this limitation, the information contained within diagnostic reports can be used as weak labels to train predictive algorithms for WSI classification [4]. Beyond image classification, the extraction of structured

SEBD 2023: 31st Symposium on Advanced Database System, July 02–05, 2023, Galzignano Terme, Padua, Italy

*Corresponding author.

laura.menotti@unipd.it (L. Menotti); stefano.marchesin@unipd.it (S. Marchesin); gianmaria.silvello@unipd.it (G. Silvello)

https://www.dei.unipd.it/~menottilau/ (L. Menotti); https://www.dei.unipd.it/~marches1/ (S. Marchesin); https://www.dei.unipd.it/~silvello/ (G. Silvello)

0000-0002-0676-682X (L. Menotti); 0000-0003-0362-5893 (S. Marchesin); 0000-0003-4970-4554 (G. Silvello)

knowledge from diagnostic reports can empower several downstream tasks; such as, clinical decision support [5], keyword search [6], visual analytics [7], and automated annotation support [8].

The significant increase in the volume of clinical data can benefit from Information Extraction (IE) techniques, helping to reduce the burden of manual data curation [9, 10]. In this regard, clinical IE has been applied to a variety of clinical text formats, such as radiology reports [11], discharge summaries [12], and pathology diagnoses [13, 14]. Rules are still common in clinical IE systems as an integration tool for Machine Learning (ML) models – resulting in hybrid systems. These methods leverage the strengths of both ML and rule-based architectures to achieve high performance [14]. In [7], the Semantic Knowledge Extractor Tool (SKET) has been presented. SKET is an unsupervised hybrid knowledge extraction system that extracts entities from text and links them to concepts in a reference ontology; namely, the ExaMode ontology [15]. SKET has been designed to extract knowledge from diagnostic reports for three types of cancer: colon carcinoma, uterine cervix cancer, and lung cancer.

In this work, we adapt SKET to a newer, improved version of the ExaMode ontology and we extend the approach to an additional use case: Celiac disease. The new version of SKET is publicly available at https://github.com/ExaNLP/sket. Due to the nature of Celiac disease, we broaden the scope of SKET to not only identify the presence of specific concepts, but also extract additional information related to them. Furthermore, we introduce some consistency checks that ensure the identified concepts are compliant with what occurs in practice. We perform an experimental evaluation to assess the effectiveness of the new version of SKET. Overall, the new version obtains an average performance gain of 8.17% compared to the old one. On the other hand, the new SKET version reaches an accuracy of 0.9484 on Celiac disease.

The rest of the paper is organized as follows. Section 2 outlines the source data while Section 3 describes the ExaMode ontology. Section 4 summarizes the system architecture and presents the new version of SKET. Section 5 provides the evaluation setup and reports the effects of the changes on the previous use cases, together with results on the Celiac use case. Finally, Section 6 concludes the paper.

2. Source Data

The development and evaluation of SKET involved diagnostic reports from two European medical centers, namely the Azienda Ospedaliera per l'Emergenza Cannizzaro (AOEC) located in Catania, Italy, and the Radboud University Medical Center (RUMC) situated in Nijmegen, The Netherlands. Diagnostic reports incorporate the findings of pathology tests and follow the College of American Pathologists (CAP) international guidelines[1] for pathology reports [16, 17]. AOEC provided pathology reports written in Italian for all four use cases. On the other hand, RUMC reports are written in Dutch and comprise reports for colon and cervix cancer, as well as Celiac disease cases. In this case, reports were produced using speech-to-text tools, making them more verbose with respect to AOEC, where reports are collected in the clinical workflow. Compared to the source data used in [7], the two medical centers provided additional reports for the three cancer use cases, and supplied 2,576 reports about Celiac disease. In particular, the

[1] https://www.cap.org/protocols-and-guidelines

two medical centers provided 4,016 additional colon reports, 7,017 new reports about uterine cervix cancer, and 235 additional reports concerning lung cancer.

Most of state-of-the-art Named Entity Recognition (NER) and Entity Linking (EL) methods over unstructured data are in English. Thus, following [7], we translate Italian and Dutch reports in English using the open-source, pre-trained Marian Neural Machine Translation (NMT) models [18] and sanitize translated text from common errors through the use of handcrafted rules.

3. The ExaMode Ontology

The ExaMode ontology is a multi-lingual resource conceived to encode digital histopathology diagnostic reports associated with WSIs. It was designed by analyzing textual records and following an iterative process with continuous feedback and validation from pathologists and clinicians. Specifically, the ExaMode ontology defines the relevant concepts and properties organized into five semantic areas concerning clinical case reports (i.e., general aspects), diagnosis results, performed tests, interventions employed to retrieve the specimen, and the anatomical location of the findings. The new version of the ExaMode ontology[2] preserves the original structure but we heavily revised the Celiac disease part while focusing on providing a consistent ontology, where all elements comply with a set of design standards. To this end, we established some principles to reference external taxonomies to limit the creation of new classes to a minimum and reuse existing ontologies as much as possible. In addition, we determined a design model exploiting the Simple Knowledge Organization System (SKOS) data model to apply to classes whose individuals are general concepts. Overall, we revised the previous use cases to ensure a more comprehensive representation of diagnostic reports in histopathology.

For what concerns cancer-related use cases, the total number of elements is almost identical between the two versions. Nevertheless, we removed some elements used in the previous version of the ontology and we included some new ones. For instance, for colon cancer, we introduced 6 new elements and removed 4 old ones. In particular, we removed some broader concepts in favor of specific ones, e.g., different degrees of dysplasia. This results in a wider spectrum of concepts to be matched in the EL component of SKET. We applied the same methodology also for cervix and lung use cases, where we added some elements about koilocytes and Human Papilloma Virus (HPV), for cervix cancer, and one class to identify lung carcinoma findings, for the lung use case.

Most of the changes to the ExaMode ontology involve the Celiac disease use case, which has undergone a complete redesign to more accurately reflect the domain of interest. Due to the availability of diagnostic reports about Celiac disease, we were able to conduct an in-depth analysis of the domain and validate the ExaMode ontology. As a result, we integrated 37 new elements to account for the heterogeneity of symptoms and findings associated with Celiac disease. In particular, we added several intestinal abnormalities and findings such as malabsorption, gastric metaplasia, and erosion, together with information about villi that can be useful when diagnosing Celiac disease. For instance, we added elements about the villi length, their absence, their degree of atrophy, and the presence of flattened ones. We also added two

[2]The new version of the ExaMode ontology is available at http://examode.dei.unipd.it/ontology/.

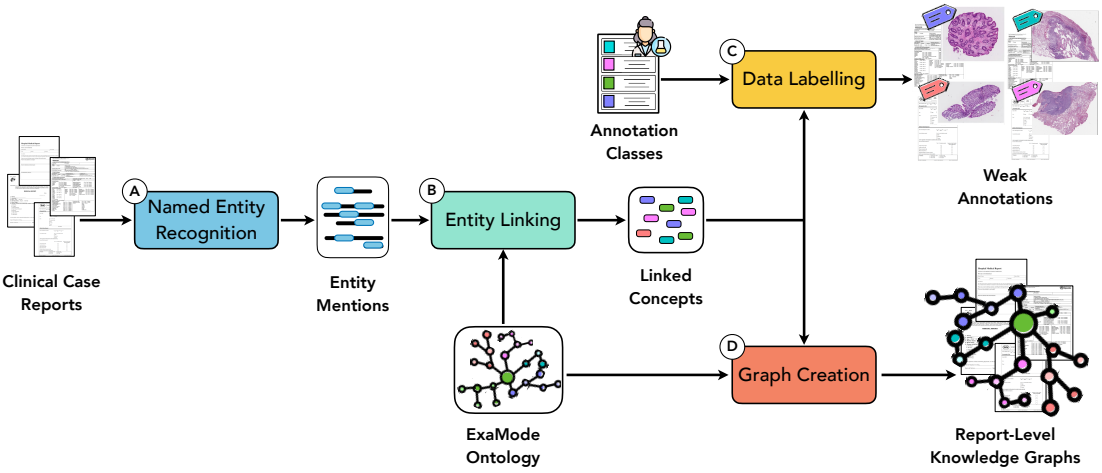

Figure 1: SKET architecture. We extract relevant concepts from clinical case reports and link them to the ExaMode ontology. The main modules are: (A) Named Entity Recognition, (B) Entity Linking, (C) Data Labelling, and (D) Graph Creation. Linked concepts can be used as weak labels for weakly supervised tasks or as nodes to build report-level knowledge graphs.

data properties to specify the severity of the duodenitis and the stage of Celiac disease based on the classification system proposed by Marsh-Oberhuber [19, 20].

4. SKET 2.0

In this work, we adapt SKET to the updated ExaMode ontology by aligning some aspects of the NER and EL modules and by tailoring the Graph Creation module to the new data schema. We recall that in [7] only three use cases were considered; namely, colon carcinoma, uterine cervix cancer, and lung cancer. In this work, we extend SKET to Celiac disease. In this regard, we expand each module to account for the new use case. In particular, we add Celiac-related rules to the NER and EL modules, and we also introduce a mapping between Celiac concepts and annotation classes.

4.1. System Architecture

SKET adapts pre-trained NER models and employs unsupervised EL methods to extract relevant concepts from diagnostic reports and link them to the ExaMode ontology. Extracted information can serve as weak labels to train predictive models for image classification tasks [4] or as nodes to build knowledge graphs based on the ontology data schema. We report the system's architecture in Figure 4.1. We preserve the same architecture of [7] but we adapt each module implementation to the new version of the ontology and to a novel use case concerning Celiac Disease. SKET consists of 4 modules: (A) Named Entity Recognition, (B) Entity Linking, (C) Data Labelling, and (D) Graph Creation. Note that components (A) and (B) are sequential, while (C) and (D) can run in parallel.

The *Named Entity Recognition* module identifies entities within the clinical case report text. SKET employs a hybrid-NER system, where ScispaCy models [21] and Neural Language Models [22, 23] are combined with hand-crafted rules to refine the outputs. Rules have been developed by analyzing and identifying common behaviors among a set of diagnostic reports and are available in the SKET GitHub repository[3].

In the *Entity Linking* module, extracted entities are linked to the ExaMode ontology components. SKET solves the EL task by introducing a two-stage model, where similarity matching is employed when rule-based, ad-hoc matching fails. SKET also presents a post-processing step where mentions that are commonly linked to unrelated ontology concepts are removed.

The *Data Labelling* component produces annotations by mapping extracted concepts to a list of annotation classes. Through this component, SKET outputs weak labels that can be used to perform weakly supervised classification tasks [4]. For each use case, pathologists have been consulted to define the most clinically relevant set of annotation classes. The Celiac disease annotation classes are: (1) Celiac disease; (2) Non-specific duodenitis; (3) Normal.

In the *Graph Creation* component, extracted concepts are used as nodes to build report-level knowledge graphs in Resource Description Framework (RDF) format. In particular, RDF graphs are created by following the data schema provided by the ExaMode ontology. In this way, SKET can enhance the semantic understanding of the diagnostic reports [24].

4.2. Celiac disease use case

As opposed to the other use cases, Celiac disease is a non-cancerous disease and can manifest itself in a large variety of symptoms [25]. The diagnosis of Celiac disease is based on the description of the small intestine alterations, usually detected with a duodenal biopsy, by expert pathologists [26]. Microscopic analysis of duodenal samples for Celiac disease provides information about villi, enterocytes, intra-epithelial lymphocytic infiltrate, and glandular crypts. The absence or alteration of these structures is crucial for the diagnosis. Furthermore, biopsies include characteristics that have to be well described, with particular attention to increased intraepithelial T lymphocytes, decreased enterocyte height, crypt hyperplasia, and villous atrophy. As a result, we must include some data properties to encompass key aspects related to intestinal mucosa alterations.

In [7], SKET extracts mentions from the report text and links them to relevant concepts in the ExaMode ontology. However, this approach merely provides information on the presence or absence of some features valuable for the diagnosis. Thus, we broaden the scope of SKET by not only identifying the presence of specific concepts but also extracting additional information related to them. For instance, if a report includes “moderate villi atrophy" with this new approach we are able to identify the presence of the concept “villi atrophy" together with its severity, i.e., “moderate". In order to achieve this, we analyzed a restricted set of Celiac reports to identify common patterns in word phrases referring to intestinal abnormalities and diagnoses. In particular, when we identify mentions related to concepts modeled as data properties in the ontology, we employ some rule-based techniques to identify the data property values and append such information to the linked concepts. Such additional facts are then exploited in the graph creation component to instantiate the corresponding data properties.

[3] https://github.com/ExaNLP/sket/tree/main/sket/nerd/rules/

Table 1
Test data size. For each medical center, we report the number of diagnostic reports that have been manually annotated by experts. Symbol "–" indicates that the medical center did not share data about the specific use case.

	AOEC	RUMC
Colon	1,704	2,065
Cervix	1,777	2,350
Lung	1,902	-
Celiac	1,654	922

In [7], the data labeling module performs a multi-label task, allowing for multiple annotations on a single report. For example, let us consider the cervix cancer use case, one report can comprise both "Presence of HPV infection" and "Cancer - adenocarcinoma in situ" annotation classes. Conversely, in the Celiac disease use case, we assume there can only be one correct label for each report. We recall that the Celiac disease use case comprises three labels: "Celiac disease", "Non-specific duodenitis" and "Normal". Hence, the nature of the labels better suits a multi-class classification scenario. To adapt SKET for this, we include some consistency checks in the data labeling component that assess the annotation quality of Celiac reports. For cancer use cases, if the EL component does not extract any concept from the input text, SKET labels the report as "No Cancer" or "Non-informative". For Celiac disease, instead, we make SKET check whether multiple annotation classes have been identified for the same report. In that case, the output violates the multi-class assumption, so SKET removes all annotations and labels the report as "Inconclusive". Note that the label "Inconclusive" serves to reflect those cases where it is not possible to match one of the labels defined by pathologists.

5. Evaluation

5.1. Experimental Setup

We perform two evaluations: (1) we compare the two versions of SKET on cancer-related use cases using the setup defined in [7]; and (2) we evaluate the new version of SKET on the newly introduced Celiac disease use case. Table 1 reports the number of annotated reports for each use case and medical center. For Celiac Disease, the test dataset comprises 456 reports labeled "Celiac Disease", 102 reports concerning "Non-specific duodenitis" and 2,018 normal reports. Note that the label distribution across the dataset is imbalanced due to the fact that we are relying on data coming from a real-case scenario. In other words, certain conditions occur more often than others in the clinical routine.

5.2. Experimental Results

5.2.1. Cancer-related use cases

To perform the first evaluation, we employ the same set of manually labeled reports as in [7] and compare results with those reported in [7] for the original use cases. Table 2 reports the

Table 2
Data labeling results on colon, cervix, and lung cancer pathology reports. The considered measures are subset accuracy and weighted F1. We perform a comparison between the results reported in [7] (column "[7]") and this work (column "This Work").

Use-case	Accuracy		Micro F1		Weighted F1	
	[7]	This Work	[7]	This Work	[7]	This Work
Colon	0.7525	0.8644 (↑ 14.87%)	0.8386	0.9139 (↑ 8.98%)	0.8373	0.9139 (↑ 9.15%)
Cervix	0.5281	0.5537 (↑ 4.85%)	0.7791	0.8049 (↑ 3.31%)	0.7611	0.8199 (↑ 7.73%)
Lung	0.8137	0.8528 (↑ 4.81%)	0.8387	0.8907 (↑ 6.20%)	0.8262	0.8873 (↑ 7.39%)

results obtained by the new version of SKET for the three cancer use cases. Concerning results, we notice that the new version of SKET achieves better results in all of them for all measures, with an average accuracy gain of 8.17%. The performance boost for each use case reflects the changes on the SKET architecture as well as the updates to the ExaMode ontology, which are described in Section 4. Indeed, since the colon cancer use case experienced the highest amount of updates in the ontology, it exhibits a peak gain of 14.87% in terms of accuracy. These results show that SKET performances are affected by the reference ontology used in the EL module. Specifically, the ExaMode ontology defines the set of concepts to be matched in the report text. This, in turn, affects the Data Labeling module which relies on the presence or absence of specific concepts to generate labels. For this reason, having a representative ontology is a crucial aspect to ensure good performance.

5.2.2. Celiac disease use case

To assess the quality of the labels extracted by SKET concerning the newly-added use case, diagnostic reports presented in Table 1 have been manually labeled by experts.

Table 3 reports the results on data labeling for the use case concerning Celiac disease. We report the performance of two runs to assess the effect of the consistency checks and post-processing step: "Base" and "Full". "Base" refers to the standard configuration of SKET, without any additional steps. On the other hand, "Full" comprises both the post-processing step and consistency checks. Results confirm the advantages of adding these two modules since an improvement has been generated from eliminating inaccurately matched concepts and labels. Concerning the "Full" run, SKET achieves the highest performance scores among all use cases, with an accuracy of almost 0.95. This can be attributed to the smaller number of labels, i.e. 3, and the multi-class classification assumption. Moreover, Celiac reports usually follow a more structured format. Thus, the language ambiguities – typical of free text – that hinder IE applications are limited. Results demonstrate SKET is effective on Celiac Disease, confirming the adaptation power of the system architecture to unseen scenarios.

Table 4 reports the performance distribution across the different annotation classes for Celiac disease. In particular, for all labels, we evaluate precision, recall, and F1 score. Results demonstrate that the new version of SKET is able to correctly predict all labels with high performance. In particular, precision is above 0.9 for all classes while recall and F1-score are between 0.82 and 0.97. These oscillations can be attributed to the lower number of annotated

Table 3
Data labeling results on Celiac disease use case. For each run, we report the accuracy, micro F1, and weighted F1 scores. Each run refers to different system settings.

Run	Accuracy	Micro F1	Weighted F1
Base	0.9289	0.9368	0.9382
Full	**0.9484**	**0.9536**	**0.9531**

Table 4
Data labeling results on Celiac disease use case. For each annotation class, we report the precision, recall, and F1 score.

Annotation Class	Precision	Recall	F1-score
Celiac Disease	0.90	0.86	0.88
Non-specific duodenitis	0.90	0.82	0.86
Normal	0.97	0.98	0.97

reports for “Celiac Disease" and “Non-specific duodenitis" with respect to normal reports.

6. Conclusion

In this work, we adapt SKET to the new version of the ExaMode ontology and we extend it to a novel use case concerning Celiac disease. We preserve the same architecture presented in [7], but we revise the NER and EL modules and we tailor the Graph Creation module to build report-level knowledge graphs based on the updated data schema. Concerning Celiac disease reports, we implement each module to address the novel use case. Due to the nature of diagnostic reports about Celiac disease, we broaden the scope of SKET by not only identifying the presence of specific concepts but also extracting additional information related to them to populate relevant data properties. Our results outperform [7] in all original use cases in all 3 performance measures, with an average accuracy gain of 8.17%. The performance boost for each use case reflects the changes made to SKET as well as the updates to the ExaMode ontology. Concerning the Celiac disease use case, the system achieves the highest performance scores among all use cases, with an accuracy of 0.9484 and a weighted F1 score of 0.9531. The performance on the novel use case demonstrates that SKET is effective on Celiac disease and confirms its ability to adapt to new, unseen scenarios.

Acknowledgments

This work was supported by the ExaMode Project, as a part of the European Union Horizon 2020 Program under grant 825292.

References

[1] S. S. Cross, Underwood's Pathology: A Clinical Approach, Elsevier Health Sciences, 2013.

[2] M. G. Hanna, A. Parwani, S. J. Sirintrapun, Whole slide imaging: Technology and applications, Advances in Anatomic Pathology 27 (2020). URL: https://journals.lww.com/anatomicpathology/Fulltext/2020/07000/Whole_Slide_Imaging__Technology_and_Applications.5.aspx.

[3] E. A. Krupinski, A. R. Graham, R. S. Weinstein, Characterizing the development of visual search expertise in pathology residents viewing whole slide images, Human pathology 44 (2013) 357–364.

[4] N. Marini, S. Marchesin, S. Otálora, M. Wodzinski, A. Caputo, M. Van Rijthoven, W. Aswolinskiy, J.-M. Bokhorst, D. Podareanu, E. Petters, S. Boytcheva, G. Buttafuoco, S. Vatrano, F. Fraggetta, J. Van der Laak, M. Agosti, F. Ciompi, G. Silvello, H. Müller, M. Atzori, Unleashing the potential of digital pathology data by training computer-aided diagnosis models without human annotations, npj Digital Medicine (2022) 102. URL: https://doi.org/10.1038/s41746-022-00635-4. doi:https://doi.org/10.1038/s41746-022-00635-4.

[5] A. P. Glaser, B. J. Jordan, J. Cohen, A. Desai, P. Silberman, J. J. Meeks, Automated extraction of grade, stage, and quality information from transurethral resection of bladder tumor pathology reports using natural language processing, JCO Clinical Cancer Informatics (2018) 1–8. URL: https://doi.org/10.1200/CCI.17.00128. doi:10.1200/CCI.17.00128. arXiv:https://doi.org/10.1200/CCI.17.00128, pMID: 30652586.

[6] A. Badan, L. Benvegnù, M. Biasetton, G. Bonato, A. Brighente, A. Cenzato, P. Ceron, G. Cogato, S. Marchesin, A. Minetto, L. Pellegrina, A. Purpura, R. Simionato, N. Soleti, M. Tessarotto, A. Tonon, F. Vendramin, N. Ferro, Towards Open-Source Shared Implementations of Keyword-Based Access Systems to Relational Data, in: Proc. of the Workshops of the EDBT/ICDT 2017 Joint Conference (EDBT/ICDT 2017), Venice, Italy, March 21-24, 2017, volume 1810 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017. URL: https://ceur-ws.org/Vol-1810/KARS_paper_01.pdf.

[7] S. Marchesin, F. Giachelle, N. Marini, M. Atzori, S. Boytcheva, G. Buttafuoco, F. Ciompi, G. M. Di Nunzio, F. Fraggetta, O. Irrera, H. Müller, T. Primov, S. Vatrano, G. Silvello, Empowering digital pathology applications through explainable knowledge extraction tools, Journal of Pathology Informatics 13 (2022) 100139. doi:https://doi.org/10.1016/j.jpi.2022.100139.

[8] F. Giachelle, O. Irrera, G. Silvello, Medtag: a portable and customizable annotation tool for biomedical documents, BMC Medical Informatics Decis. Mak. 21 (2021) 352.

[9] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. F. Jones, R. Forshee, M. Walderhaug, T. Botsis, Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review, Journal of Biomedical Informatics 73 (2017) 14–29. URL: https://www.sciencedirect.com/science/article/pii/S1532046417301685. doi:https://doi.org/10.1016/j.jbi.2017.07.012.

[10] S. Marchesin, G. Silvello, TBGA: a large-scale gene-disease association dataset for biomedical relation extraction, BMC Bioinform. 23 (2022) 111.

[11] S. Hassanpour, C. P. Langlotz, Information extraction from multi-institutional radiology reports, Artificial Intelligence in Medicine 66 (2016) 29–39. URL: https://

www.sciencedirect.com/science/article/pii/S0933365715001244. doi:https://doi.org/10.1016/j.artmed.2015.09.007.

[12] H. Yang, I. Spasic, J. A. Keane, G. Nenadic, A text mining approach to the prediction of disease status from clinical discharge summaries, Journal of the American Medical Informatics Association 16 (2009) 596–600. URL: https://www.sciencedirect.com/science/article/pii/S1067502709000929. doi:https://doi.org/10.1197/jamia.M3096.

[13] N. Ashish, L. Dahm, C. Boicey, University of california, irvine–pathology extraction pipeline: The pathology extraction pipeline for information extraction from pathology reports, Health informatics journal 20 (2014) 288–305.

[14] E. Santus, T. Schuster, A. M. Tahmasebi, C. Li, A. Yala, C. R. Lanahan, P. Prinsen, S. F. Thompson, S. Coons, L. Mynderse, R. Barzilay, K. Hughes, Exploiting rules to enhance machine learning in extracting information from multi-institutional prostate pathology reports, JCO Clinical Cancer Informatics (2020) 865–874. URL: https://doi.org/10.1200/CCI.20.00028. doi:10.1200/CCI.20.00028. arXiv:https://doi.org/10.1200/CCI.20.00028, pMID: 33006906.

[15] S. Marchesin, L. Menotti, G. Silvello, The examode ontology (full version, v.2) (2.0) [data set], 2023. doi:https://doi.org/10.5281/zenodo.7669237.

[16] J. R. Srigley, T. McGowan, A. MacLean, M. Raby, J. Ross, S. Kramer, C. Sawka, Standardized synoptic cancer pathology reporting: A population-based approach, Journal of surgical oncology 99 (2009) 517–524.

[17] D. Ellis, J. Srigley, Does standardised structured reporting contribute to quality in diagnostic pathology? the importance of evidence-based datasets, Virchows Archiv 468 (2016) 51–59.

[18] M. Junczys-Dowmunt, R. Grundkiewicz, T. Dwojak, H. Hoang, K. Heafield, T. Neckermann, F. Seide, U. Germann, A. F. Aji, N. Bogoychev, A. F. T. Martins, A. Birch, Marian: Fast neural machine translation in C++, in: Proceedings of ACL 2018, System Demonstrations, Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 116–121. URL: https://aclanthology.org/P18-4020. doi:10.18653/v1/P18-4020.

[19] M. N. Marsh, Gluten, major histocompatibility complex, and the small intestine: a molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'), Gastroenterology 102 (1992) 330–354.

[20] G. Oberhuber, G. Granditsch, H. Vogelsang, The histopathology of coeliac disease: time for a standardized report scheme for pathologists., European journal of gastroenterology & hepatology 11 (1999) 1185–1194.

[21] M. Neumann, D. King, I. Beltagy, W. Ammar, ScispaCy: Fast and robust models for biomedical natural language processing, in: Proceedings of the 18th BioNLP Workshop and Shared Task, Association for Computational Linguistics, Florence, Italy, 2019, pp. 319–327. URL: https://aclanthology.org/W19-5034. doi:10.18653/v1/W19-5034.

[22] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in neural information processing systems 26 (2013).

[23] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational

Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[24] M. Agosti, S. Marchesin, G. Silvello, Learning unsupervised knowledge-enhanced representations to reduce the semantic gap in information retrieval, ACM Trans. Inf. Syst. 38 (2020). URL: https://doi.org/10.1145/3417996. doi:10.1145/3417996.

[25] V. Villanacci, A. Vanoli, G. Leoncini, G. Arpa, T. Salviato, L. R. Bonetti, C. Baronchelli, L. Saragoni, P. Parente, Celiac disease: histology-differential diagnosis-complications. a practical approach, Pathologica 112 (2020) 186.

[26] V. Villanacci, P. Ceppa, E. https://www.overleaf.com/project/617c1835b0e93b5f11026f98Tavani, C. Vindigni, U. Volta, On behalf of the "Gruppo Italiano Patologi Apparato Digerente (GIPAD)" and of the "Società Italiana di Anatomia Patologica e Citopatologia Diagnostica"/International Academy of Pathology, Italian division (SIAPEC/IAP) Coeliac disease: the histology report, Digestive and Liver Disease 43 (2011) S385–S395. doi:10.1016/S1590-8658(11)60594-X.