“D2.2”

Conceptual descriptive framework for multimodal knowledge

Version: 1.3

Last Update: 22/12/21

Distribution Level: *PU*

Distribution level

PU = Public,

RE = Restricted to a group of the specified Consortium,

PP = Restricted to other program participants (including Commission Services),

CO= Confidential, only for members of the ExaMode Consortium (including the Commission Services)

Project supported by European Union

Horizon 2020 grant agreement 825292

The ExaMode Project Consortium groups the following Organizations:

Partner Name	**Short name**	**Country**
HAUTE ECOLE SPECIALISEE DE SUISSE OCCIDENTALE	HES-SO	Switzerland
UNIVERSITA DEGLI STUDI DI PADOVA	UNIPD	Italy
SIRMA AI	ONTOTEXT	Bulgaria
STICHTING KATHOLIEKE UNIVERSITEIT	Radboudumc	Netherlands
MICROSCOPEIT SP ZOO	MICROSCOPEIT	Poland
AZIENDA OSPEDALIERA PER L'EMERGENZA CANNIZZARO	AOEC	Italy
SURFSARA BV	SURFSARA BV	Netherlands

Document Identity

Creation Date:	08/11/2021
Last Update:	22/12/2021

Revision History

Version	**Edition**	**Author(s)**	**Date**
0	1	Svetla Boytcheva	08/11/2021
Comments:	Initial draft version		
0	2	Pavlin Gyurov	15/11/2021
Comments:	Applications description		
0	3	Nikola Tulechki	24/11/2021
Comments:	Semantic model description		
0	4	Pavlin Gyurov, Nikola Tulechki, Todor Primov	06/12/2021
Comments:	Contributions to the first draft of the deliverable		
1	0	Svetla Boytcheva	20/12/2021
Comments:	Completion of the first version of the deliverable		
1	1	Gianmaria Silvello	21/12/2021
Comments:	Review comments		
1	2	Todor Primov	21/12/2021
Comments:	Reflection on the review comments		

1	3	Svetla Boytcheva	22/12/2021
Comments:	The final version of the deliverable		

Executive summary

This deliverable presents the designed conceptual descriptive multimodal knowledge framework based on the Examode data, combining extracted information and metadata from textual data (synoptic reports) and Whole Slide Images (WSIs). This will allow medical experts to easily represent into a unified framework multimodal medical information, reducing the effort to combine textual and imaging data for clinical decision support systems. The deliverable contains a short description and of different methods for multimodal knowledge graph modeling.

The core component of the designed conceptual descriptive multimodal knowledge framework is the semantic model described in Histographer (D6.1 "Requirement Analysis for Image Centric Semantic Diagnostics Knowledge Management System", D6.2 " Software Architecture and Specification for Semantic Diagnostics Knowledge Management System" and D6.3 "Adaptable Semantic Diagnostics Knowledge Management Prototype"). HistoGrapher knowledge graph (KG) contains integrated patient data from Electronic Health Records (EHRs) and synoptic reports and mappings with various domains specific ontologies, as well as the specially designed Examode ontology. The semantic model of Histographer is extended with additional concepts and relations representing the WSIs metadata and the information related to the WSIs annotations. The proposed conceptual descriptive multimodal knowledge framework addresses all types of Examode data described in D3.1 "First set of curated, publicly available multimodal and multimedia data" and D3.5 "Final set of cured publicly available multimodal and multimedia data".

There are presented some applications of the designed conceptual descriptive multimodal knowledge framework for faceted and similarity search that could be used as an assistance service for a clinical decision support system.

Table of Contents

Table of Figures

Index of Tables

List of abbreviations

API	Application Programming Interface
CSV	Comma Separated Values
EHR	Electronic Health Records
ETL	Extract Transform Load
ICD-10	International Classification of Diseases, 10th revision
ICD-O	International Classification of Diseases, Oncology
HIS	Hospital Information System
HTTP	Hypertext Transfer Protocol
KG	Knowledge Graph
KGE	Knowledge Graph Embedding
LIMS	Lab Information Management System
NLP	Natural Language Processing
RDF	Resource Description Framework
REST	Representational state transfer
SNOMED CT	SNOMED Clinical Terms
TCGA	The GDC Cancer Genome Atlas
TSV	Tab-Separated-Values
UMLS	Unified Medical Language System
VS	Vector Space
WSI	Whole Slide Image

1 Introduction

The main objective of the designed conceptual descriptive multimodal knowledge framework is to enable the investigation of deeper insights into the knowledge presented by different types of modalities, including structured and unstructured data. The dataset includes – images (WSIs), raw text (synoptical reports), structured data (EHRs), extracted knowledge from the text (Histographer pipelines), annotated WSIs (metadata).

The potential applications of the multimodal knowledge framework will be as a model for the decision support systems, allowing a search for similar cases and serving as an assistance tool to the pathologist during the diagnostic process.

The designed model is based on the data described in the D3.5 "Title: Final set of cured publicly available multimodal and multimedia data".

As a core component of the designed conceptual descriptive multimodal knowledge framework is the semantic model described in Histographer (D6.1 "Requirement Analysis for Image Centric Semantic Diagnostics Knowledge Management System", D6.2 " Software Architecture and Specification for Semantic Diagnostics Knowledge Management System" and D6.3 "Adaptable Semantic Diagnostics Knowledge Management Prototype").

The current semantic model of the Histographer is based on textual data only and needs some extension and adaptation in order to address the extracted knowledge from the WSIs.

The process includes the following steps:

- analysis of all metadata for the WSIs,
- mappings of the metadata for different image datasets BERN, RUMC, AOEC.
- Identification of the approaches for a combination of the image and textual information.
- Designing of methods for visualization of the image information in the knowledge graph (KG).
- Definition of methods for similarity search in the multimodal KG

2 Types of Multimodal Knowledge Modelling

The term "multimodal knowledge" has various interpretations addressing the problem of integration of knowledge extracted from sources with different modalities like images, text, audio, video, etc.

Very early models include separate processing pipelines for each type of modality and merging techniques. Another approach is cross-modality translation and unification to a single modality model of all different modalities. The more advanced approaches try to address all modalities together and to use a joint model for knowledge presentation and processing.

Table 1 Types of Multimodal Knowledge Modelling

version	Description	Pros	Cons
1	Individual pipelines for text and images that are merged in KG [1]	Experience, Available Components	Non-beneficiary from the information from images and text
2	Separate pre-trained DL models for text and images, that are merged in a single VSM (vector space model) [1] [4] [9]	Available pre-trained image model Available pre-trained language model - bioBERT[1] [10], ClinicalBERT[2] [11]	Additional data are necessary for the interpretation and integration. Same cases to be processed by both models
3	Single DL model trained on annotated text and images [2] [3] [5] [6]		Computationally expensive – needs a significant amount of data to train such a model.
4	Version 1 + KGE on the top of the generated KG [7]	Experience, Available Components	Sparse knowledge graph, Computationally expensive - need many training data
5	Multimodal Faceted Search on KG + Similarity Search on KGE - Individual search for textual and image data, presenting merged results (Information Retrieval Approach) [8]	Experience, Available Components	

Some aspects of the possible design of the conceptual descriptive multimodal knowledge framework are listed in the table below with a sketch of some major issues regarding their

[1] https://github.com/dmis-lab/biobert
[2] https://github.com/EmilyAlsentzer/clinicalBERT

implementation for the Examode scope. References to the publications with a more detailed description of these approaches are included for each version.

After the analyses of all advantages and disadvantages of the possible solutions, version 5 was selected which needs a significantly smaller amount of annotated data and the model can be easily scaled and extended for new data and new use cases.

3 Conceptual descriptive framework for multimodal knowledge

The objective of this framework is to ensure homogeneous access to structured and unstructured multimodal data stored in the knowledge graph. Thus its multimodality is a matter of integration between the different data sources of the Histogrpaher platform[3] in a real clinical setup. In order to achieve this objective, we will extend the Examode Ontology[4] so that visual data can be fully supported.

The current version of Histogrpaher[5] contains various ontologies (ICD-10[6], ICD-O[7], UMLS[8], SNOMED[9], Human Disease Ontology[10], PathLex[11], Mondo[12], ProstateCancer[13]) alongside Examode Ontology as well as patient data from AOEC reports.

Based on the images provided in the BERN, and RUMC datasets (see D3.1) we combine the metadata supplied by the BERN and RUMC WSIs.

[3] https://marketplace.big-data-value.eu/content/histographer
[4] http://examode.dei.unipd.it/ontology/
[5] http://examode.ontotext.com/
[6] https://icd.who.int/browse10/2010/en#/
[7] https://apps.who.int/iris/bitstream/handle/10665/96612/9789241548496_eng.pdf
[8] https://www.nlm.nih.gov/research/umls/index.html
[9] https://www.snomed.org/
[10] https://disease-ontology.org/
[11] https://bioportal.bioontology.org/ontologies/PATHLEX
[12] https://mondo.monarchinitiative.org/
[13] https://bioportal.bioontology.org/ontologies/PCAO

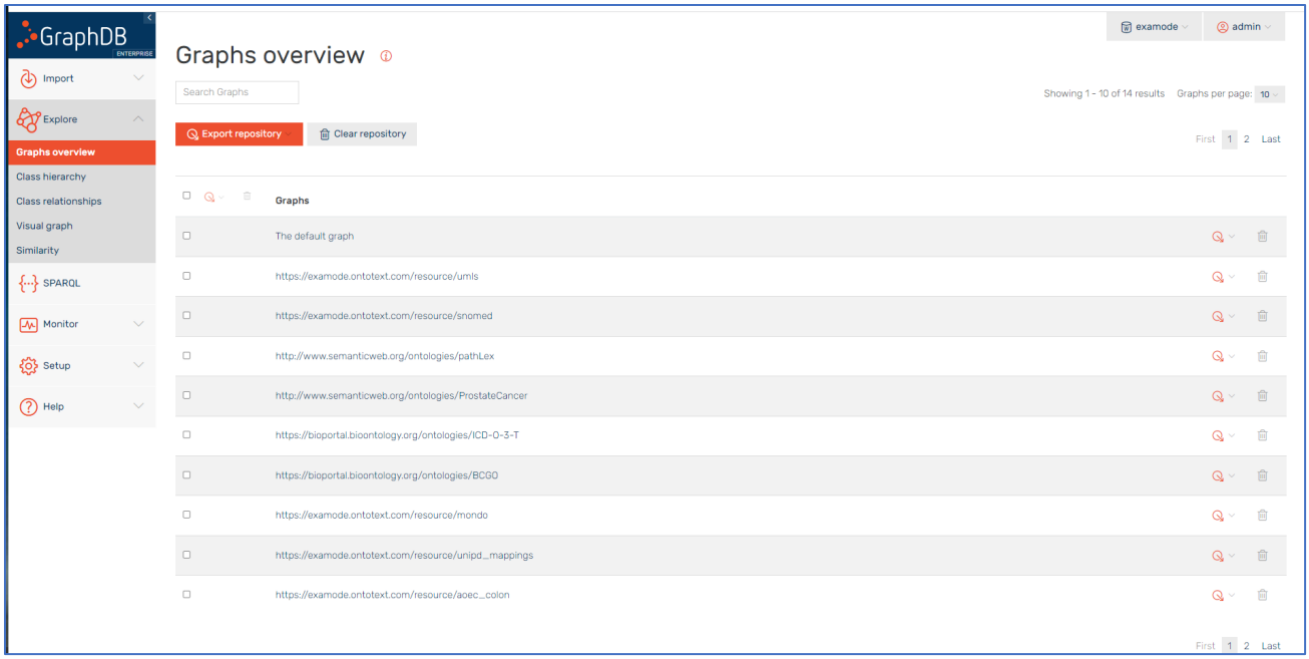

Figure 1 Histographer resources

3.1 Data

The proposed conceptual descriptive multimodal knowledge framework addresses all types of Examode data described in D3.1 “First set of curated, publicly available multimodal and multimedia data” and D3.5 “Final set of cured publicly available multimodal and multimedia data”.

Three different data sources can be combined in the semantic model:

- Textual descriptions of performed examinations, localization of the examined tissue, some remarks, and patient data like age and gender.
- WSIs metadata.
- Histopathological image analysis annotations on the slide images.

3.2 Extensions of the Semantic Model

Currently, each clinical case is represented as a knowledge subgraph in the whole knowledge graph holding all the semantic data in the system. Different search, grouping and similarity techniques are applied so that the relevance and similarity between clinical cases can be assessed on the basis of shared properties and their values - explicitly provided or derived. The annotations are generated using the semantic data model and annotation scheme which assures their compatibility and comparability among the KG.

Slides are provided as stored in MIRAX format[14] (.mrxs). We transform the metadata from the **`Slidedat.ini`** file in the **mirx** bundle to RDF so that the values are accessible in the integrated knowledge graph and linked to the nodes representing the relevant case report.

The **`SlideDat.ini`** file contains a set of key-value pairs with various information from basic metadata like the date and hardware of the slide to more specific data such as the zoom level or various configuration options of the camera recording the particular slide.

This piece of data (Fig. 1) produces the following RDF. Note that the set of attributes corresponds to the keys in the **ini** file, thus allowing simple transformation between structures.

[14] https://openslide.org/formats/mirax/

```
[GENERAL]
SLIDE_VERSION = 01.03
SLIDE_NAME = DigitalSlide_B1M_1S_1
PROJECT_NAME = ProjectName
SLIDE_ID = D31697A372D847538311202D6F280F1A
IMAGENUMBER_X = 784
IMAGENUMBER_Y = 1136
SLIDE_CREATION_FINISHED = True
CURRENT_SLIDE_VERSION = 2.2
IMAGE_OVERLAP_MICROMETERS_X = 0
IMAGE_OVERLAP_MICROMETERS_Y = 0
SLIDE_POSITION_X = 0
SLIDE_POSITION_Y = 0
SLIDE_CONTENT = DIGITAL_SLIDE
SLIDE_CREATIONDATETIME = 06/06/2019 16:36:40
VIMSLIDE_CAMERA_REAL_BITDEPTH = 8
CONFOCAL = NO
DISK_POSITION = 0
FOCUS_MAP = AUTO
FOCUS_LIMIT = DISABLED
FLAT_FIELD_CORRECTION = DISABLED
EXTENDED_FOCUS_ALGORITHM = FL Image
VIMSLIDE_SLIDE_BITDEPTH = 8
CAMERA_TYPE = Adimec Q-12A-180Fc
ADAPTER_SIZE = 1.6
OPTOVAR_SIZE = 1
OBJECTIVE_MAGNIFICATION = 20
OBJECTIVE_NAME = Plan-Apochromat
SLIDE_TYPE = SLIDE_TYPE_BRIGHTFIELD
CameraImageDivisionsPerSide = 8
```

Figure 2 Example of the GENERAL section of the SlideDat.ini file

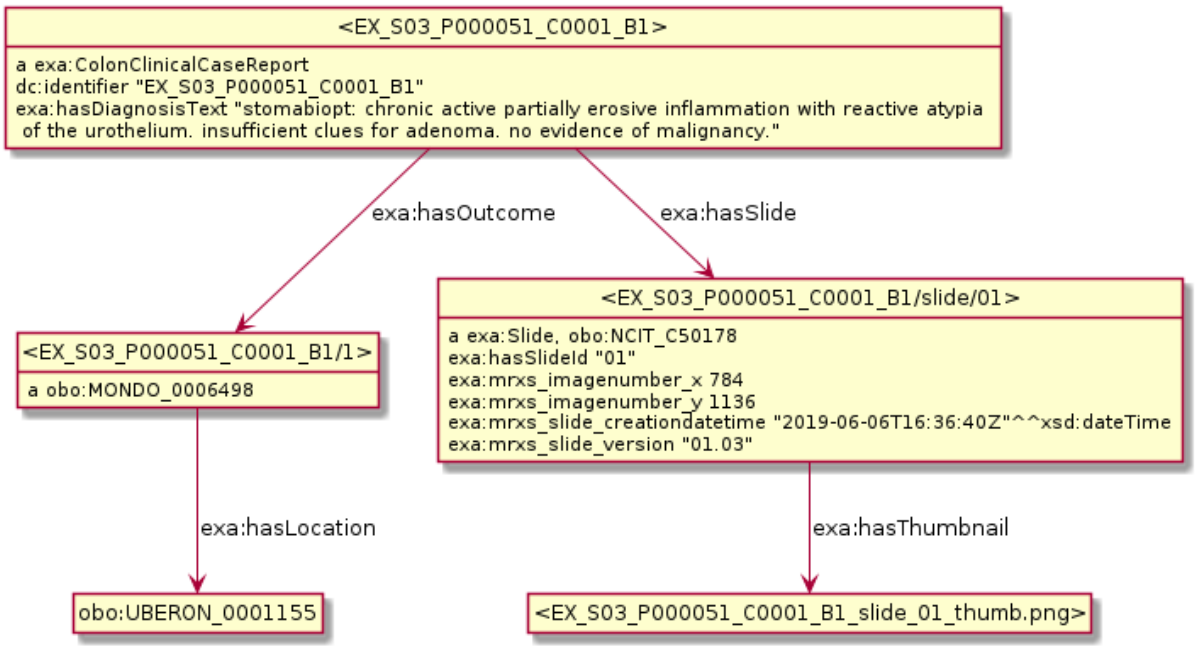

Figure 3 RDF subgraph of a Clinical Case report with a slide and thumbnail image

Given that the actual MIRAX image data can be difficult to process, thumbnails are stored using the **`exa:hasThumbnail`** attribute of the **`exa:Slide`** object.

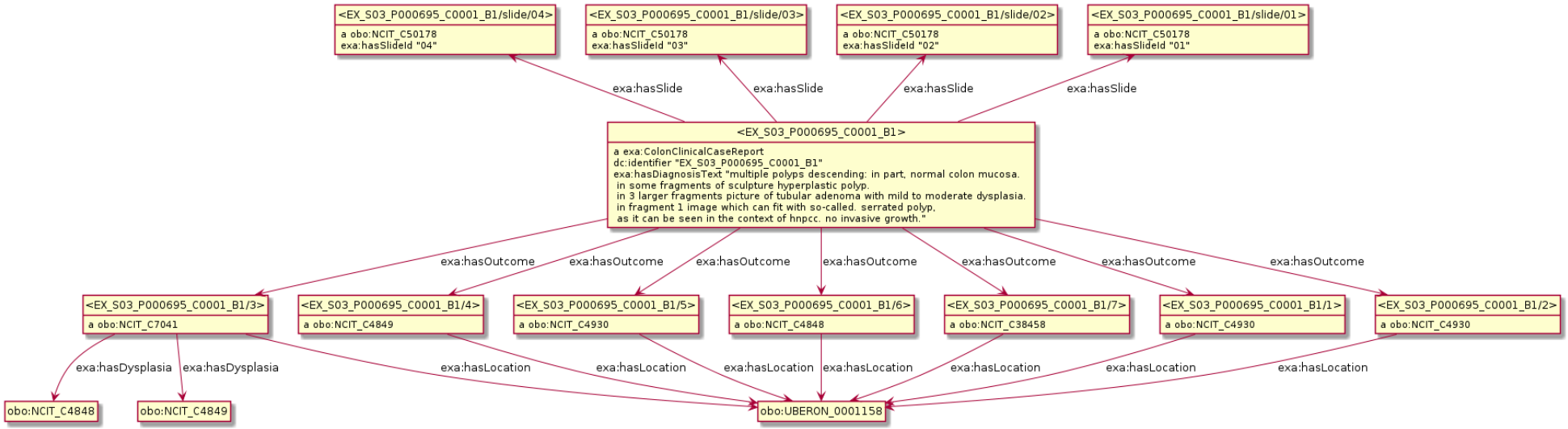

Figure 4 RDF subgraph of a Clinical Case report with multiple WSIs and multiple outcomes

4 Applications

Each of the above data sources has its own processing pipeline. However, all the pipelines share a common clinical case identifier in order for their results to be linked to each other.

A detailed description of the input pipelines and data can be provided after the finalization of the specification of integration points. It makes sense only in the context of some clinical environment setup where it should be integrated into the existing HIS or LIMS. It is also necessary that some additional simple functionality of LIMS be implemented so that integration with that system can happen.

Therefore, a short conceptual description of the Examode application view of the multimodality is provided and the specific details are omitted, or options are proposed.

The textual input data are processed by the Synopses ETL described in D6.2 and D6.3. The input is provided as data records of specific structure: rows of Microsoft Excel (xls, xlsx), Comma separated values (CSV) or Tab-Separated-Values (TSV) files with predefined field names. At the time being, they are provided manually by storing the files in a specific shared input folder. Development of REST application entry point is provisioned so that data records to be ingested in the system as rest calls. There is also an integration of some translation software provisioned which allows the internal annotation schemes and NLP processes to be implemented only in English but allows the processing of multilingual input. Currently, the options with implemented interfaces to them are Google Cloud Translation API[15] and EasyNMT[16]. Anyway, the system architecture allows their replacement by other modules, since they are separate independent components.

WSIs are usually of huge sizes which do not imply their storage several times. In the ExaMode ontology they are represented as **`exa:Slide`** objects and connected to the case reports using the **`exa:hasSlide`** predicate.

Since they are not directly used in the multimodal graph they do not need to be stored in the semantic storage nor any other side storage of the Histographer. Instead, links to their accessible storage places (probably in the LIMS) could be provided and included in the KG as values of the property **`exa:hasThumbnail`**. on the **`exa:Slide`** object These values are to be used only for visualization purposes so that it makes sense for image thumbnails to be used instead of whole size images. These links could be provided together with the textual synopses data or by a separate HTTP POST request (still not implemented) providing also the unique identifier of the clinical case.

Additionally, there is a (small capacity) file storage in the Histogrpaher. It is currently used for storing icons and other system-specific files. Optionally, it is possible that storage to be used

[15] https://console.cloud.google.com/apis/api/translate.googleapis.com/overview?pli=1&project=snappy-cistern-202114

[16] https://github.com/UKPLab/EasyNMT

also for storing thumbnails and they to be provided in the place of links within the HTTP requests, but this option should be additionally negotiated with the hospital.

The image analysis results as per currently shared information with VIRTUM (D5.6: "Prototype of product - VIRTUM-P2") consist of arrays of polygons with their annotations using SNOMED CT codes. The latter could also have some metrics like confidence, the number of occurrences, sizes of polygons, or some other metrics allowing comparison of the slides. It is almost unlikely to compare the regions themselves (e.g by evaluating and comparing their shapes and mutual collocations) to be of some additional benefit. However, their metadata – like their sizes, surfaces, numbers, etc. especially after some normalization could provide additional clues about image similarity. However, this is only a negotiation point and needs the relevant histopathological domain expert assistance in order to be defined. Here we just point out that such an approach is feasible.

Currently, there is no direct integration between MICROSCOPEIT and Sirma AI about the data transfer of these annotations and metadata and their integration in the KG. The semantic integration of the normalized textual data (clinical synopsis) and the annotated images will be done through a single point of integration with the LIMS systems in the hospital environment. The annotated images associated with the medical records will be submitted to LIMS, from which The HistoGrapher application can retrieve both unstructured textual clinical synopsis data that will be processed and normalized, but also the annotations of the whole slide images associated with the clinical case. This is scheduled for the next version of the software application. One of the possible options for integration is for these image processing annotations to be supplied by LIMS– as separate calls or combined ones. However, this is still to be specified at the time this document is written.

For now, we assume that there is a mechanism and functionality for the before-mentioned image analysis results to be provided. When that happens, they have to be converted to RDF using the Examode multimodal data model and stored in GraphDB[17] where they become a part of the KG. This functionality is provisioned in the architectural delivery D6.2, High-level Architecture diagram as Graphical Analysis results from ETL. The full integration between MICROSCOPEIT and Sirma AI is planned to be finalized for M48 and will be described in the D6.4 "Validation of Semantic Diagnostics Knowledge Management Prototype in Hospital Setting".

4.1 Visualization

Image analysis results and provided textual data about a clinical case become a part of the KG after their semantic processing and mapping. At this point further they are included in search results. There are two main ways of data visualization in the Histographer:

- grid data visualization where the corresponding data fields are displayed in a tabular grid allowing additional exploring of linked concepts

[17] https://www.ontotext.com/products/graphdb/

- Visual graph - Interactive visualization where each entity is displayed as a node in a graph, and the properties are directed arcs. The properties themselves can have additional characteristics like confidence forming so-called RDF-star[18].

RDF-Star (also known as "RDF*") allows descriptions to be added to edges in a graph such as scores, weights, temporal aspects, and provenance to edges in a graph. Formally, RDF* extends the RDF graph model by allowing statements about statements, i.e., one can attach metadata, which describe an edge in a graph, while RDF allows statements to be made only about nodes.

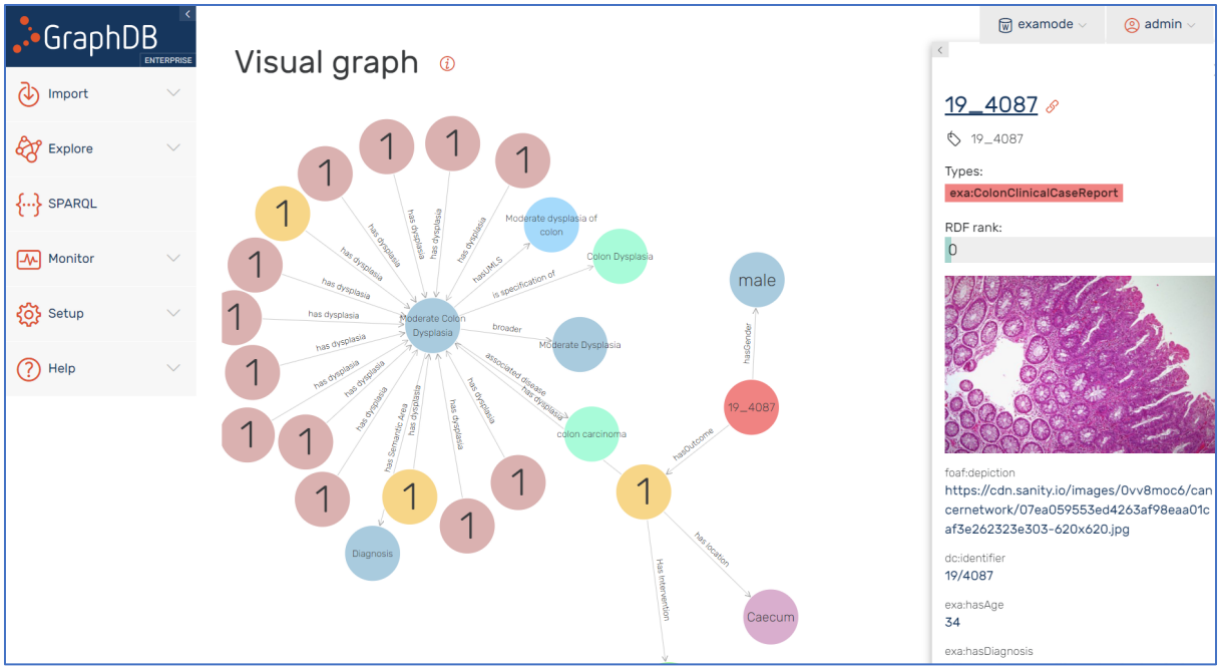

Figure 5 Multimodal knowledge graph with visualization of both structured clinical information (ExaMode Ontology and clinical instance data representing clinical cases) and associate histopathology image data

Both visualization methods are described in D6.3 Adaptable Semantic Diagnostics Knowledge Management Prototype so that no more details are provided here.

The additional point related to the multimodality is the visualization of the case thumbnails. If the below requirements are met the image is displayed in the grid of the query results as a separate field:

- ✓ The predefined query gathers the `schema:hasImage` image property as ?image field in its result set
- ✓ The corresponding entity (object) has its image link in the KG,
- ✓ the link points to an image
- ✓ the link is active
- ✓ the user has permission to download the image from the link

[18] https://www.w3.org/2021/12/rdf-star.html

4.2 Multimodal Knowledge Search

Combining data from different sources in a single place without reduction gives more opportunities for better and more specific search query definitions resulting in better matches and thus more benefits for the users of the system.

It is worth mentioning that despite the exact category matches using the Examode ontology the matches of the numerical properties like age, number of artifact occurrences, image metrics, etc. cannot be done directly, because the numbers 20 and 21 are as different as 20 and 50 for instance. That's why intervals are used instead of distinct values and values are mapped to their corresponding intervals before searching and similarity calculation.

One such mapping is done the rest of the search process uses the tooling for the execution of predefined parametric queries provided with GraphDB as well as tools for faceted and similarity search.

4.3 Similarity Search

The current state-of-the-art similarity calculation and link prediction algorithms use pre-calculated semantic indices and only objects included in that indices are searchable. That means that new indexing must be performed after ingestion of new data and before using the similarity search. There is no way for the new data to be added and the current index to be amended.

The above results in the limitation that there must be three separate steps – adding new data, indexing, and searching.

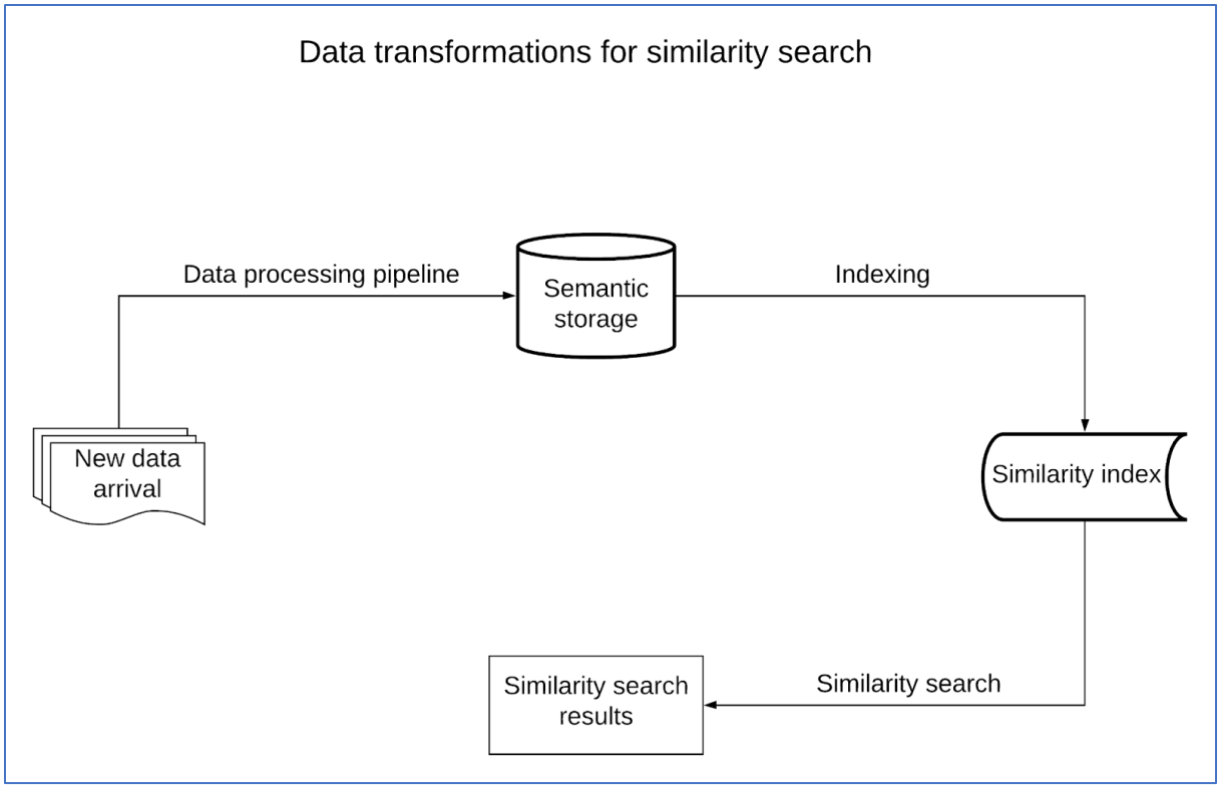

Figure 6 The pipeline for similarity search

In the held project meetings we discussed with the project members involved in the hospital environment evaluation a workflow which consists of entering batches of new data once a day and they are exploring on the next day. So, the re-indexing step being quite fast for offline processing will not produce any delays and inconsistency.

At the time being the Histographer uses Vector Space (VS) similarity index (Predication-based semantic indexing)[19] [12]. There were discussions about implementing the Knowledge Graph Embedding (KGE) approach and research on which exactly KGE algorithm to be used. However, all KGE algorithms require a significant initial amount of labeled data because the embeddings are calculated as a result of the Machine Learning step, which in its turn has these data requirements. Currently, we do not have enough data for the application of KGE.

On the other hand, the VS Index, implemented in GraphDB performs sufficiently well on small to medium amounts of data and is the best choice option. The VS index needs to be pre-computed in order to query efficiently. As a reference, the indexing of the whole Pubmed[20] abstracts collection (> 30 million abstracts) takes between 6 and 8 hours on a virtual machine with 32 GB of RAM and 4-8 CPUs. Once computed the query for similar documents and or concepts takes a few seconds.

[19] https://graphdb.ontotext.com/documentation/standard/semantic-similarity-searches.html
[20] https://pubmed.ncbi.nlm.nih.gov/

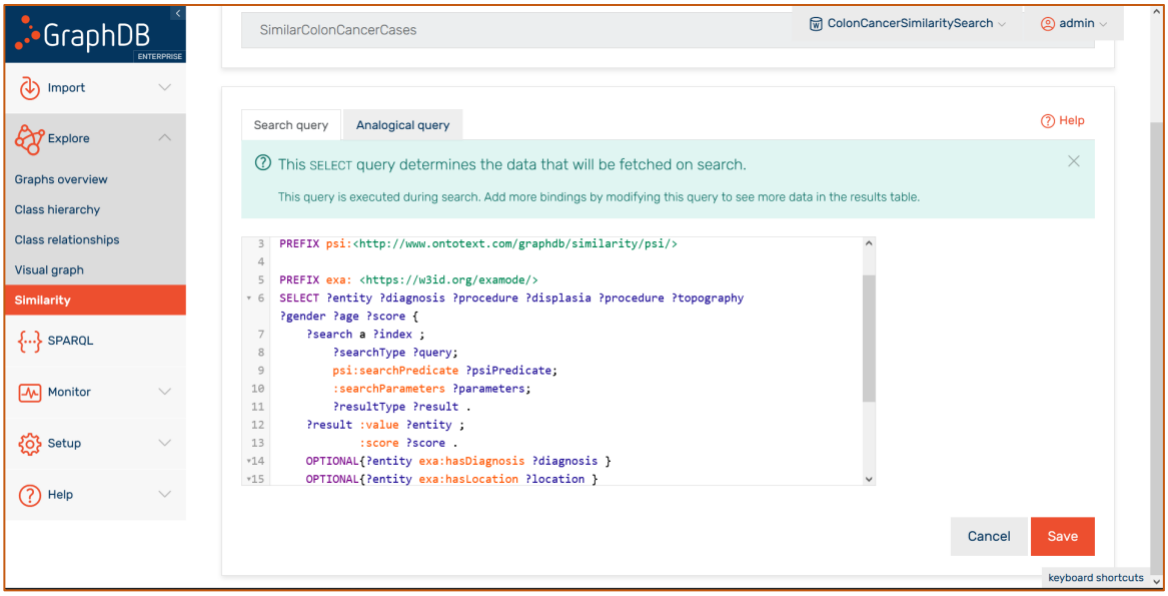

Figure 7 Configuration of the Similarity Search Index – SPARQL query for searching analogical cases in the multimodal knowledge graph

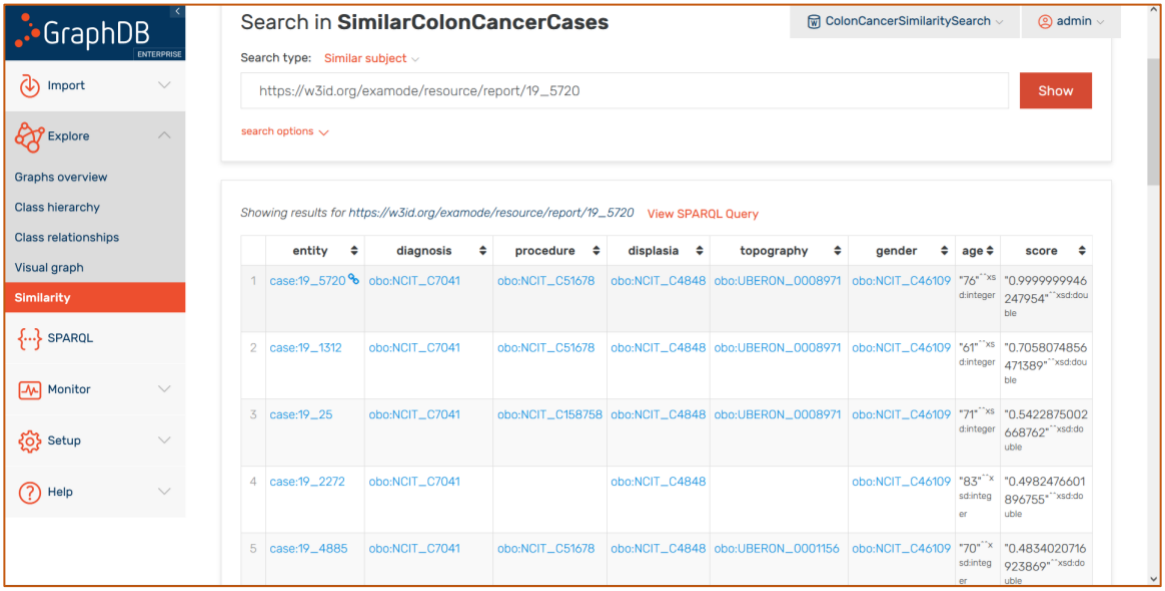

Figure 8 Similarity search for other cases related to information for patient ID 19_5720

5 Conclusion

The proposed model for multimodal knowledge graph allows to integrate the extracted knowledge from different modalities, to explore different modalities through faceted search, and in real-time to explore different knowledge embeddings in the vector space model using similarity search. The main advantages of the proposed model are that it is light, scalable, flexible and it can be easily extended for new data. It does not require a huge amount of

annotated data for training. The similarity search always uses all actually available data in the KG and does not need some pretraining and further extensions and updates of the VS model. Thus when more patient information is available always it will be taken into consideration in the decision support system.

6 References

[1]. Chang, Shih-Fu. "Multimodal Knowledge Graphs." https://aihealth.ischool.utexas.edu/IWKG_KDD2020/slides/Shih-Fu.pdf

[2]. Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011, January). Multimodal deep learning. In *ICML*. https://people.csail.mit.edu/khosla/papers/icml2011_ngiam.pdf

[3]. Liu, K., Li, Y., Xu, N., & Natarajan, P. (2018). Learn to combine modalities in multimodal deep learning. *arXiv preprint arXiv:1805.11730*. https://arxiv.org/pdf/1805.11730.pdf

[4]. Xue, Z., Ren, S., Gao, Z., & Zhao, H. (2021). Multimodal Knowledge Expansion. *arXiv preprint arXiv:2103.14431*. https://arxiv.org/pdf/2103.14431.pdf

[5]. CLIP: Connecting Text and Images https://openai.com/blog/clip/

[6]. Open-sourcing PyTorch-BigGraph for faster embeddings of extremely large graphs https://ai.facebook.com/blog/open-sourcing-pytorch-biggraph-for-faster-embeddings-of-extremely-large-graphs/

[7]. Wilcke, X., Bloem, P., de Boer, V., & van't Veer, R. End-to-End Learning on Multimodal Knowledge Graphs. http://www.semantic-web-journal.net/system/files/swj2727.pdf

[8]. Kannan, A. V., Fradkin, D., Akrotirianakis, I., Kulahcioglu, T., Canedo, A., Roy, A., ... & Al Faruque, M. A. (2020, October). Multimodal knowledge graph for deep learning papers and code. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 3417-3420). https://dl.acm.org/doi/10.1145/3340531.3417439

[9]. Mousselly-Sergieh, H., Botschen, T., Gurevych, I., & Roth, S. (2018, June). A multimodal translation-based approach for knowledge graph representation learning. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics (pp. 225-234). https://aclanthology.org/S18-2027.pdf

[10]. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, *36*(4), 1234-1240. https://arxiv.org/abs/1901.08746

[11]. Alsentzer, E., Murphy, J. R., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. *arXiv preprint arXiv:1904.03323*. https://arxiv.org/pdf/1904.03323.pdf

[12]. Cohen, T., Schvaneveldt, R. W., & Rindflesch, T. C. (2009). Predication-based semantic indexing: Permutations as a means to encode predications in semantic space. In AMIA Annual Symposium Proceedings (Vol. 2009, p. 114). American Medical Informatics Association. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815384/