A Multimodal Search Engine for Medical Imaging Studies

Eduardo Pinho[1] **· Tiago Godinho**[1] **· Frederico Valente**[1] **· Carlos Costa**[1]

Published online: 25 August 2016

Abstract The use of digital medical imaging systems in healthcare institutions has increased significantly, and the large amounts of data in these systems have led to the conception of powerful support tools: recent studies on content-based image retrieval (CBIR) and multimodal information retrieval in the field hold great potential in decision support, as well as for addressing multiple challenges in healthcare systems, such as computer-aided diagnosis (CAD). However, the subject is still under heavy research, and very few solutions have become part of Picture Archiving and Communication Systems (PACS) in hospitals and clinics. This paper proposes an extensible platform for multimodal medical image retrieval, integrated in an open-source PACS software with profile-based CBIR capabilities. In this article, we detail a technical approach to the problem by describing its main architecture and each sub-component, as well as the available web interfaces and the multimodal query techniques applied. Finally, we assess our implementation of the engine with computational performance benchmarks.

Keywords Content-based image retrieval · Computer systems · Graphical user interface (GUI) · Information storage and retrieval · PACS · Reproducibility of results · Software design · Multimodal information retrieval · Query fusion · Web services

Introduction

The use of digital medical imaging systems in healthcare institutions has increased significantly, becoming a valuable tool for medical diagnosis, decision support, and treatment procedures. Research and industry efforts to develop medical imaging equipment, including new acquisition modalities and information systems, are intense and have been grounded by the wide acceptance of the Picture Archiving and Communication System (PACS) concept. The number of medical imaging studies is constantly growing, resulting in tremendous large amounts of data produced. It is estimated that the USA will produce over 1 exabyte (=1000 petabytes = 1 million terabytes) of imaging data in 2016 [1]. Digital Imaging and Communications in Medicine (DICOM) is the standard used for storage and exchange of structured medical imaging data. A persistent DICOM object may include numerous data elements, such as pixel data, meta-data, and reports [2].

One of the most important advantages of using PACS is the facilitated sharing; seamless access to medical data; and orchestration of distinct hardware, services, and personnel, including from mobile platforms [3]. However, medical imaging repositories are often looked upon as inert bags of imaging objects that are accessible only through the DICOM query and retrieve service, using a limited number of search attributes (e.g., patient name, study ID, procedure date). Nevertheless, the means by which we currently search for information have been shaped by search engine interfaces, and free text searching is a common feature expected from any information system. Typing on a search bar with keywords or phrases of

✉ Eduardo Pinho
eduardopinho@ua.pt

Tiago Godinho
tmgodinho@ua.pt

Frederico Valente
fmvalente@ua.pt

Carlos Costa
carlos.costa@ua.pt

[1] DETI/IEETA, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

interest, although very common, is not the only way of obtaining useful information. Further advancements have granted the ability to search using pictures, audio, and other kinds of multimedia content as part of the query. Indeed, the continued efforts in image processing, medical informatics, and information retrieval are creating suitable conditions for the integration of multimodal information retrieval in the workflows of clinicians, lecturers, and researchers [4, 5]. Multimodality in the generic context of information retrieval refers to the theories, algorithms, systems, and challenges of indexing and retrieving multiple modes (kinds) of data, which may include meta-data, free text, images, or other multimedia sources [6]. In the sub-field of medical imaging systems, we apply this definition to the available information in medical imaging repositories. Therefore, the scope of retrieval covers medical image meta-data, pixel data, vital signs, structured reports, and other annotations.

Content-based medical image retrieval (CBMIR) holds great potential in medical applications by allowing the system to determine the level of similarity with existing images, which may translate to similar clinical cases [5]. These systems, however, should not rely on visual similarity alone: the combination of medical image feature sets with non-visual data, such as DICOM meta-data and structured reports, is proven to be highly beneficial for medical decision support systems [4], as the fusion of multiple modalities can provide complementary information and increase the accuracy of the overall decision-making process [7]. Visual feature extraction and comparison techniques are already contemplated by many content-based image retrieval (CBIR) systems, but these capabilities are only the tip of the iceberg. The next generation of systems should enhance these features with richer image descriptions, in order to translate them to semantic concepts, thus providing a system for *medical case* retrieval, instead of *image* retrieval. Therefore, studying the concept of multimodality in medical imaging informatics, as well as new and better ways to employ it, is a great step towards this goal. The application of multimodal information retrieval in the context of PACS is a challenging issue to be addressed and requires the development of new ways to store, index, process, and retrieve medical information. In fact, the usage of multimodal retrieval tools is extremely rare in clinical practice and is still mostly limited to the scope of research.

This paper presents a platform for making multimodal searches in a medical imaging repository, supporting complex queries composed by the combination of textual and visual information. The proof of concept was developed using Dicoogle, an open-source PACS archive, and an existing CBIR platform [8], thus withholding existing contributions in a decoupled and modular fashion. The result is a highly flexible architecture for executing multimodal searches in a PACS, for benchmarking and for clinical use, as well as a web-based platform that addresses functionality and usability concerns.

Related Work

Quite often, the architecture for multimodal information retrieval in the scope of medical imaging informatics is tightly coupled with the goals of CBIR: techniques for query refinement, expansion, and combination are usually presented as part of the image retrieval engine, which also contemplate text-based retrieval in some cases. Multimodal information retrieval has had its impact in a multitude of fields, and several tools and techniques for CBMIR have emerged over the last two decades [4], [9]. In [10], the authors cover the state of the art on multimodal medical information retrieval in three perspectives, one of which is the latest research done in CBMIR.

The NovaMedSearch engine [11] exhibits the similar goals of supporting multimodal queries with a simple and intuitive user interface, for medical case-based retrieval. Our work, in contrast, is not tightly coupled to specific sources of data and shows a greater concern of integrating the engine to a PACS. The Khresmoi project also stands out. It is a large EU-funded project with the goal of conceiving a multi-lingual and multimodal search and access system for biomedical information [12]. The main user interface is based on the ezDL project, but an alternate interface was developed, called Shambala [14]. Markonis et al. [15] have covered the use of Khresmoi for the retrieval of medical images in a PACS archive and the biomedical literature. The search engines developed under this project are backed by ParaDISE [13], a CBIR system featuring an architecture with scalability and extensibility in mind, although lacking in details about how the complete system is orchestrated in a typical usage. Rahman et al. [16] also present an interesting multimodal framework with an embedded hierarchical image classifier and a fixed pipeline of fusion strategies for medical image retrieval. It was our intention in this new architecture to be as flexible as possible in the kinds of queries that can be created, by supporting query trees of arbitrary depth, configurable transformation, and fusion strategies, and the possibility of including classifiers as a dedicated source of data, which do not have to rely on the system's extracted features.

The presented solution was built over Dicoogle [17], a platform-independent PACS archive that replaces the traditional relational database with a more agile indexing and retrieval mechanism. It was designed with automatic extraction, indexation, and storage of all meta-data detected in DICOM medical images, including private attribute tags, without re-engineering or reconfiguration requirements [18]. Its plugin-based software architecture (as seen in Fig. 1) allows us to add new means of extraction and storage of multiple types of information associated with the same study, without modifying

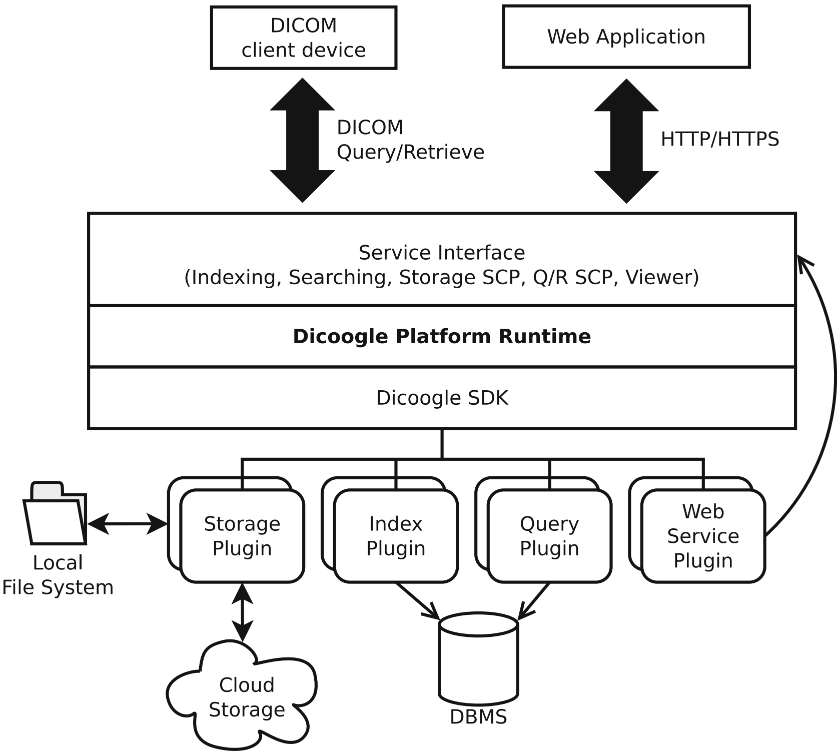

Fig. 1 General architecture of Dicoogle

the core software, hence increasing overall robustness and applying extensions to the platform at deployment time. Although the presented architecture is applicable to other systems, Dicoogle stands as an ideal platform for our implementation, given the aforementioned factors.

Dicoogle's engine has since been augmented with CBIR [8], supporting automatic image feature extraction on indexation, as well as similarity metrics for performing query-by-example. The concept of *CBIR profile* was introduced in order to cope with the rapid appearance of new feature extraction and similarity techniques that may only be compatible with the content of a certain modality. A CBIR profile contains information about the similarity metric used, the required features to successfully apply it, and the target modalities.

Methods and Materials

In order to provide multimodal search capabilities to Dicoogle, a new plugin was developed with the following main objectives:

- Create an interoperability layer among different information sources, namely text-based and image-based query providers, as well as potentially other information modalities in the future
- Integrate state-of-the-art query fusion techniques and leverage the potential of CBIR systems such as Dicoogle's CBIR plugin to be put in image retrieval benchmarking scenarios, as well as in clinical practice
- Exploit a flexible and usable search user interface relying on state-of-the-art paradigms in the field, such as query-by-example and relevance feedback

Architecture

Figure 2 presents a top-level view of the proposed multimodality search plugin and its interactions with the Dicoogle runtime framework. The plugin does not contain feature extraction, similarity metrics, or direct means of querying a database. Such tasks are delegated to existing query providers by categorization of their modality. In Dicoogle, all operations involving *storage*, *querying*, and *indexation* are immediately available via the SDK's internal API.

The two entry points for multimodal queries are the RESTful API ("API and Data Representation" section) and the user interface ("Graphical User Interface" section), which are both web-based. At the back-end, the *multimodal search engine* orchestrates the full search process, which depends on sub-procedures divided in three categories: *query transforms*, *query fusions*, and *result transforms*. The *query interface manager* contains adapters for delegating queries to specific providers, according to the kind of query requested by the engine. The characteristics and behaviors of each component are further detailed throughout the "Query Formulation and Processing Workflow" section.

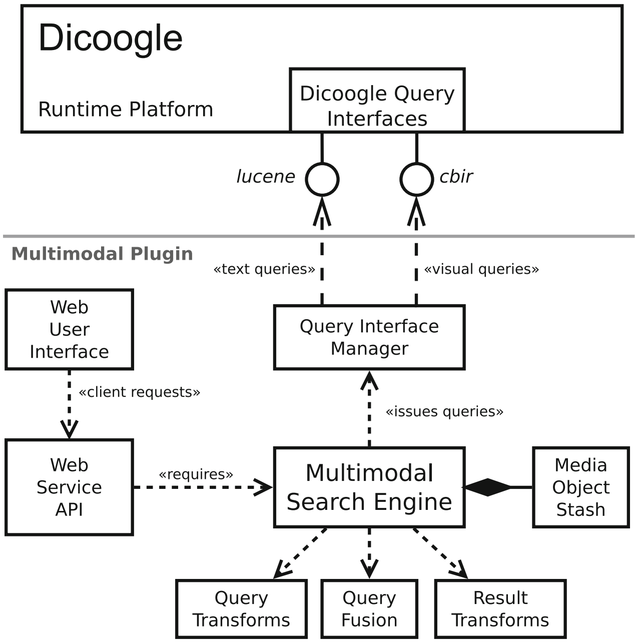

Fig. 2 Architecture diagram of the multimodality plugin (below the *gray line*), depicting its key interactions with Dicoogle. Each dashed arrow represents a dependency

The multimodal retrieval engine currently contemplates access to two modality interfaces of the Dicoogle core platform:

Textual data: Typical text queries are fundamental to support the DICOM query plugin. This interface is based on the Lucene (https://lucene.apache.org) query language, supporting both keyword-based and free text queries. An existing plugin relying on a Lucene index [18] was used as a proof of concept for a text-based query provider, but the use of other providers is still possible by converting the query to an adequate format internally.

Visual data (CBIR): Image queries rely on the query-by-example pattern and must provide either a universal resource indicator (URI) of an already indexed image, or an object containing the embedded image. These queries can optionally be followed by the name of a CBIR profile, hence focusing on a particular group of features and metrics. Additionally, as a particular form of inter-media feedback, we have contemplated a *domain filter*, which restricts similarity testing to the given set of URI-identified items. This feature is useful, as the image search domain can then be provided by a separate source.

These categories enforce a level of harmonization among plugins that accept the same kind of queries and follow the same API at the level of the Dicoogle QueryInterface [17]. In an implementation independent from Dicoogle, such APIs would have to be established from scratch.

Query Formulation and Processing Workflow

The major difference between a simple text query and a multimodal one is that it may contain information infeasible or too expensive to be fully represented in a textual format. If the user wants to perform a query-by-example over a local file, this object needs to be uploaded before or alongside the remaining description of the query. A container for temporary media content (henceforth called *media object stash*) was conceived. With this approach, multimedia files not already indexed by Dicoogle are transferred before the effective query descriptor is sent ("API and Data Representation" section). The engine will later on retrieve the object by its unique identifier and have its feature set extracted and processed by the CBIR module.

Once all required media content is stored, the multimodal search takes place according to the pipeline depicted in Fig. 3:

1. The user takes the available web-based user interface (or interacts with the system's REST API) in order to formulate and send a query.
2. The engine pre-processes the query by traversing it through a fixed series of query transformation functions. This step is where query refining is applied and additional entries are potentially included from relevance feedback.
3. The multimodal query is split into unimodal queries that are invoked on the Dicoogle core runtime, by adapting it to one or more compatible query providers. The operation yields multiple result streams.
4. The result lists are combined into a single list using late query fusion techniques, which are detailed in the "Multi-Query and Fusion Techniques" section.
5. The results may then undertake a series of transformations, such as for augmenting the results with existing DICOM attributes.
6. Before returning the final outcome, an aggregation of results can take place based on a specific level of the DICOM image model (DIM) hierarchy. Considering that the main goal points towards a case-based retrieval, *series-level* or *study-level* aggregation is essential to a practitioner's interpretation of the results.

This pipeline is established at Dicoogle deployment time and can be extended with more query/result transformers and fusion strategies.

Multi-query and Fusion Techniques

Late query fusion involves applying an algorithm over multiple ranked lists of results, with the aim of obtaining a single stream of more relevant items. They are, in general, the most preferred and utilized solutions to combining multiple searches for the past few years, and some diversity of

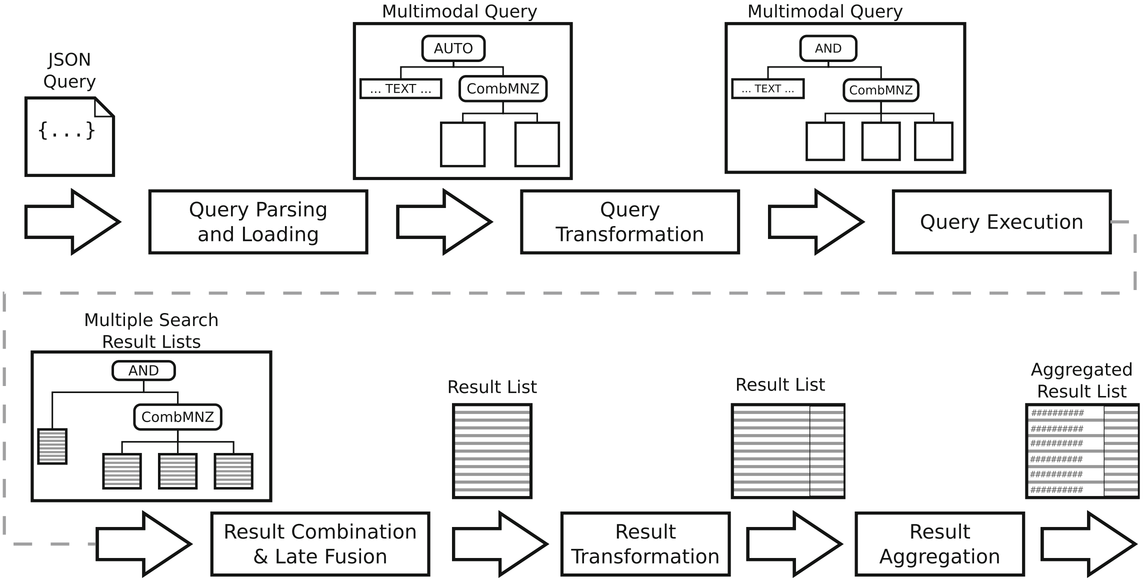

Fig. 3 Diagram depicting the system's multimodal search pipeline

algorithms in this scope have emerged. Although more complex fusion techniques based on machine learning exist [19], late fusion algorithms may be as simple as a Boolean operator applying a restriction to the result list (AND, OR), or a reordering algorithm for the combined list. The latter may be based on the score of each result entry (representing the system's individual appreciation of relevance for that result) or be based on the rank of said result on the full list. Other algorithms that do not fit in either case also exist, such as the inverted squared rank (ISR) [20].

The proposed system features multimodal queries based on a combination of multiple text and image content objects forming a tree structure of queries. A leaf of the tree represents a unimodal query, which is handled by the Dicoogle core runtime through an adapter. Any other tree node represents a query fusion process. For identification purposes, each node contains a *content object key* (CO key) property, which functions as an index for that node among its siblings in a query fusion. Thus, a path of CO keys from the root to the intended node can be used to uniquely identify a node in the multimodal query (e.g., "0.2.1" is query #1 of a query fusion, which is query #2 of the top-level query fusion).

The multiple outcomes obtained from the unimodal queries must then be combined. In this proposal, a few known late fusion algorithms are contemplated. Boolean combinations AND and OR were made to restrict a list of results, being particularly useful for combining a keyword-based condition with image queries (e.g., Modality:CT AND cbir([image])). CombSUM (Eq. 1), CombMNZ (Eq. 2), CombMAX (Eq. 3), and CombMIN (Eq. 4) were added to the initial assortment of score-based query fusion strategies, as specified in [21]. The reciprocal rank fusion (RRF) algorithm [22], which is rank-based, was also included (Eq. 5). In Eqs. 1–5, d is the document of a result; N_j is the number of sub-queries performed in the fusion; $S_j(d)$ and $R_j(d)$ are the score and rank of d in the sub-query *j*, respectively; and $F(d)$ is the number of occurrences in the sub-queries.

$$\mathrm{CombSUM}(d) = \sum_{j=1}^{N_j} S_j(d) \tag{1}$$

$$\mathrm{CombMNZ}(d) = \left(\sum_{j=1}^{N_j} S_j(d) \right) \times F(d) \tag{2}$$

$$\mathrm{CombMAX}(d) = \arg \max_{j=1:N_j} S_j(d) \tag{3}$$

$$\mathrm{CombMIN}(d) = \arg \min_{j=1:N_j} S_j(d) \tag{4}$$

$$\mathrm{RRF}(d) = \sum_{j=1}^{N_j} \frac{1}{k + R_j(d)} \tag{5}$$

Each result list may yield score values that are inadequate for comparison among different queries, since they may follow disparate score distributions and ranges [23]. Therefore, each result list needs to be normalized before a score-based fusion between other lists takes place. Rather than having a single implementation, the platform allows a client to select one out of multiple score normalization strategies. The algorithms currently included are *min-max* (proposed in [24]), *min-sum* and *min-var* (the last two proposed in [23]). This "freedom of choice" was deemed relevant due to the fact that some score normalization algorithms offer more robustness against outliers, thus increasing performance when fused by strategies that are particularly sensitive to them [23].

The search results of CBIR queries in Dicoogle have a distance-based score. That is, the value 0 represents the highest score possible, whereas higher values relate to greater dissimilarity or irrelevance among objects. This irregularity is addressed by automatically converting distance-based scores to a non-negative "*higher is more relevant*" range before each normalization.

API and Data Representation

A multimodal query representation format was specified as part of this proposal. It was designed to be simple and easy to use by web-based applications, have a low memory footprint, and support some degree of extensibility. Other means of describing multimedia queries are already available but were not as fitting for the given requirements. The Multimedia Retrieval Markup Language (MRML) [25] is a standard defining an XML-based communication protocol for performing queries to compliant multimedia retrieval systems. Although an implementation exists and the format could be extended to support multimodal queries, the protocol is unsuitable for the web, the official website[1] is no longer available at the time of writing, and the standard has not had any significant impact during the last few years. Therefore, making an additional effort to make the system MRML compliant did not seem to be worthwhile. We have established a JavaScript Object Notation (JSON) data schema for the complete description of multimodal queries. JSON has the advantage of producing files with less overhead and facilitating query construction and parsing. This is especially useful in web applications, the runtime environment of which have built-in JSON support. Other systems can also easily read and write queries with the aid of JSON libraries.

As expressed, the proposed system composes a set of web services providing 4 main resource endpoints (relative to Dicoogle's base URL for web services):

> /multimodal/search is the endpoint for performing queries. A query JSON object of the query is uploaded with a POST operation, which will follow with a response containing the outcome of the search.
> /multimodal/stash provides the means to store media objects for use in future queries. A store operation will accept either a media content (e.g., of MIME type image/png, application/dicom, …) or a multi-part data form containing the same item (MIME type multipart/form-data). This content type was added in order to support file uploads purely based on an Internet browser's implementation of HTML5.
> /multimodal/ui is used to retrieve the user interface and will be consumed by an Internet browser.
> /multimodal/fusion simply returns a list of query fusion strategies made available, as a JSON array of (value,label) pairs.

Graphical User Interface

The developed system provides a graphical user interface (Fig. 4) to explore and use query fusion techniques made available with the plugin's search engine. As a key concept of interaction, each unimodal query in the multimodal query tree is represented in a box. Empty boxes (hereby named *ghost boxes*) are shown to allow the user to introduce more queries in the tree.

The drag-and-drop paradigm was significantly exploited for this interface, as it is a simple and familiar form of interaction for the user. An image from a previous result can be dragged and dropped over a query object box in order to become part of the query. If the box was "ghosted," a new child query is contemplated and another ghost box is placed next to it. Furthermore, the user can upload an image file by dragging a file from the client's system, supporting DICOM and other generic image formats such as PNG and JPEG.

Text queries can still be performed by typing on a text input box. If the aforementioned query box already contains an image, the text input will instead provide the unimodal query's meta-options.

For a multi-query fusion, the user can drop targets in a highlighted region below a ghost box to choose one of the available fusion strategies from a drop-down list, or leave the "Automatic" option selected. Although the "automatic" strategy currently falls back to a default, this option may later on comprise a smart combination of query fusion techniques based on a query analysis.

Once issued, the results are shown as a grid of images, all of which can be dragged and dropped for a manual form of relevance feedback. Not only a grid layout of results is more fitting for drag-and-drop operations but it is also known that radiologists often prefer this layout to a list of items [26]. Any further changes in the multimodal query will automatically trigger a new search.

Computational Assessment and Results

The methods described in this document allow us to capture the immense heterogeneity of medical image archives by supporting content discovery services on top of multiple data formats. More specifically, they enable the combination of multiple providers into a single search interface where users may search for interesting artifacts using a unified query language. However, featuring such functionality alone is not enough. It is of major importance that our services are offered

[1] www.mrml.net (as seen in August 3rd 2012)

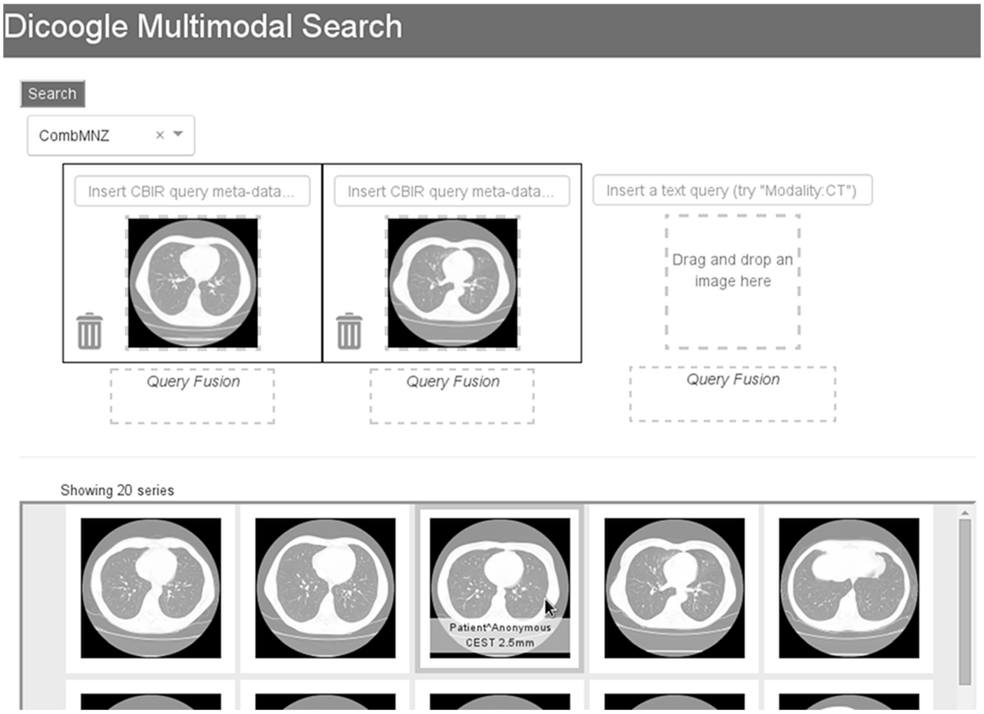
Dicoogle Multimodal Search
Search
CombMNZ
Insert CBIR query meta-data...
Insert CBIR query meta-data...
Insert a text query (try "Modality:CT")
Drag and drop an image here
Query Fusion
Query Fusion
Query Fusion
Showing 20 series
Patient^Anonymous
CEST 2.5mm

Fig. 4 A draft of the multimodal search engine's graphical user interface, depicting a multi-level query, ghost query boxes, and a few results from the search

in a performant manner according to the requirements of the medical imaging environment, which are known to be demanding. The following section presents a series of trials devised to ensure the proposed architecture's computational performance in real world medical institutions.

In this picture, scalability raises a concept of major importance. In computer science, scalability refers to a system's ability to maintain its performance indicators with increasing levels of load. The performance indicators of an interactive system, such as ours, reflect the throughput of successful requests it can handle, in our case, search requests. On the other hand, the load factors reflect the number of concurrent requests, which is an indicator of how many users are using the system at the same time, as well as their complexity. In a scalable system, the rate of degradation of the performance indicators with the increasing amount of load would be as close to zero as possible. In such a utopia, the system would be capable of handling an infinity of users simultaneously.

Taking this formal definition into consideration, we devised three experiments in order to understand the degree of scalability of proposed system. Firstly, we wanted to capture how the system responds to the complexity of the search operations. As a result, the first experiment relates the number of returned results with the time necessary to handle the search task. On the same note, the second experiment relates the latency of the search task when it is used the fusion operator. The last experiment was designed to analyze the system response in simultaneous tasks processing. The experiments were conducted using an Intel® Core™ i7-3770 CPU @ 3.40GHz × 8 with 12 GiB to run the Dicoogle instance. The dataset used for these tests was retrieved from the clinical case archive of the Belarus Tuberculosis Portal (www.tuberculosis.by). Four hundred one clinical cases were indexed, consisting of 62,198 medical images.

The first experience involved using a fixed query composed of two images and a meta-data search for the keyword "CHEST." The two images were examples of pulmonary CTs. The number of results requested in the search procedure (as configurable by the web service) varied between 1 and 1000. This query was designed merely for experimental purposes: although it makes little sense to search for 1000 related artifacts, these tests are bound to a hypothesis where this solution will scale by the number of results. In total, 3000 search operations were collected. The resulting distribution of the results can be analyzed in the scatter-plot in Fig. 5.

Empirically, it is perceptible that the increasing number of returned results has little effect on the search service time. Nevertheless, we computed a linear regression using the least squares method. The results (−~0.001 s) confirm a very slight variation rate, meaning that the number of returned artifacts has very little impact on the system's search response time. More concretely, it is expected an aggravation of 1 ms in the search time per retrieved result. In practice, this raises no concerns on its own regarding the system's scalability.

The second experiment, the results of which are presented in Fig. 6, introduced a performance comparison of the fusion operators described in this document. The goal was to discover if any of the proposed operators were impractical in a real world environment. We tested four fusion algorithms, which were best applicable to our queries. The performed queries

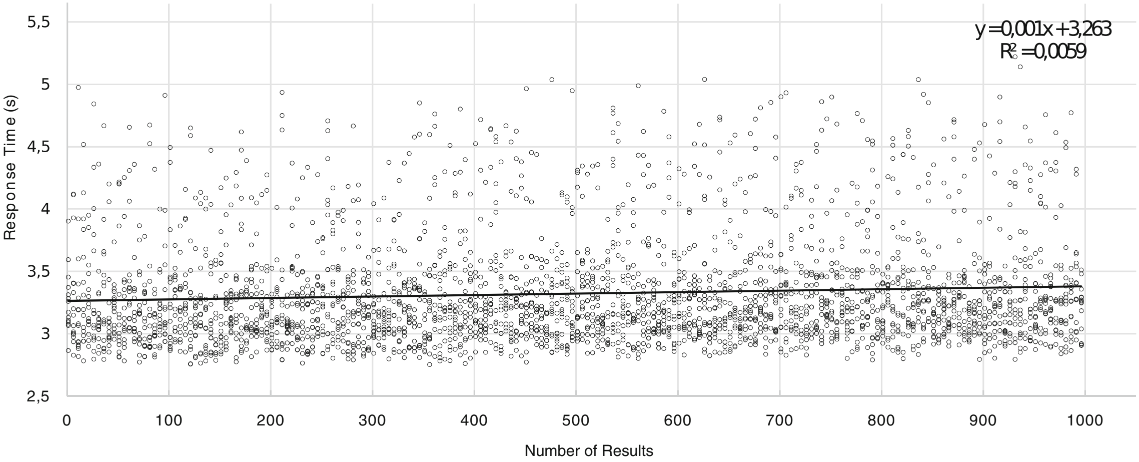

Fig. 5 Scatter-plot representing the engine's response time in terms of the number of results obtained from the query

were a combination between a meta-data query and an image query, the latter of which were filtered by an additional image query to refine the results. As expected, neither of the fusion operators proved to be impractical to use; however, we noticed that the CombSUM, CombMAX, and CombMIN operators took considerably more time than the others. This is an expected behavior, as both the intersection operator (AND) and RRF require no score normalization.

The last experiment is actually more interesting, as it evaluates the system performance in a multi-user environment. Initially, we asked an experienced user to perform six search operations over the dataset, involving late fusion operator and without any restriction of complexity. We recorded the queries inserted, but also the idle time spent by the user between each query. This capture was assumed as reasonably demonstrative of a regular usage pattern of the system. A query simulator was developed according to his pattern. The program performs the searches as a regular user, with a variable delay between queries modeled by a normal distribution ($avg = 10$ s, $std = 3$ s). The experiment consisted in running multiple instances of this program simultaneously, mimicking a regular usage of the system with a variable, but controlled, number of users. We tested up to nine users simultaneously, as we think that it is a fairly high number of concurrent users in a PACS of a central hospital.

Figure 7 shows the average cumulative response time of all searches for each user in the multiple test cases. As it is perceivable, the escalation of the cumulative response time with the increasing number of concurrent requests is best fit by a linear regression. As opposed to an exponential fitting function, a linear regression ensures that the proposed methods are easily scalable to a multitude of users, provided that adequate hardware is used. The regression also shows that a penalty of 12.5 s per concurrent request is to be expected.

Conclusions

Handling multimodal information is a multidisciplinary field. Before different kinds of data are sought to be combined, an appropriate interpretation of the underlying multimedia content is required. If the task of content-based image retrieval alone is a complex one, even more challenging will be their combination in a heterogeneous environment. A new layer of research opportunities emerges, in which such a rich amount of data may be bridged into semantic concepts.

This article proposes an architecture for multimodal information retrieval with the main objective of being usable in real world PACS scenarios, including real time search operations

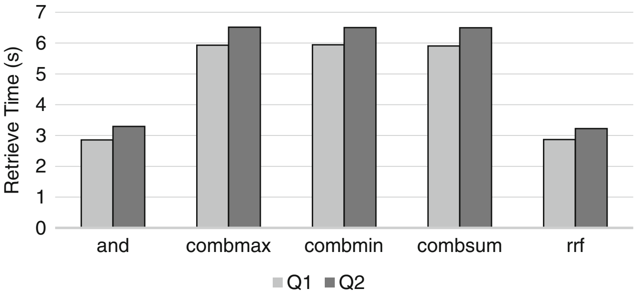

Fig. 6 Chart exhibiting the engine's response time over fusion queries with different late fusion algorithms

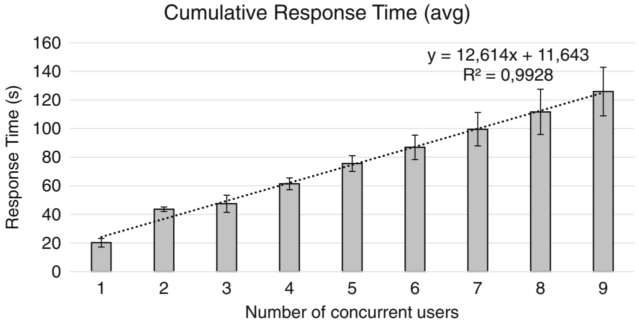

Fig. 7 Average cumulative response time measured in the third experiment

over medical imaging repositories. Its backbone was designed to be extensible, supporting new algorithms without major changes to the software, thus providing a multimodal layer of abstraction over the large domain of existing retrieval algorithms, such as feature extractors and model representations. At this level, such techniques will not be incurred a cost in retrieval quality when integrated with the platform. Rather, it allows researchers to discover improvements by combining multiple sources. A proof of concept was built as an extension to Dicoogle, although the decoupling of the architecture from this system can be done by relying on other query provider manager implementations.

Certain improvements regarding the kind-based query provider interfaces can be considered, namely how one should mark a query image as relevant or irrelevant, how regions of interest (ROI) can be outlined in an image, and how one would request a series of images to be feature-fused at the level of the original source, using which feature fusion technique (for instance, Rocchio's algorithm [27]). This extra input in the query may be very important for the integration of relevance feedback [28], which is known to disambiguate user interpretations [29]. Furthermore, these interfaces could be expanded to also return suggestions of query expansions and auto-completion when such an expansion is possible (for example, indexed DICOM tags, RadLex terms, or CBIR profile names). A complete integration of these sources will inevitably involve leveraging adequate query specification constructs into their interfaces, for use in the multimodal search engine.

It is also possible to orchestrate multiple implementations of multimodal search engines, even across institutions, as a means to attain reproducibility of retrieval results. However, the lack of an appropriate commonplace standard for performing multimodal information retrieval over the web can be a serious impediment. Both the scientific community and the industry would benefit from a long-lasting standard, the development of which makes for a promising line of research that may be based on some of the design principles of MRML [25].

The system was implemented as a web platform that addresses functionality and usability concerns, supporting complex queries composed by the combination of textual and visual information by using state-of-the-art fusion techniques. The performance and scalability of this architecture were evaluated, and the results demonstrate that the proposed solution can be used in real world environments. The system's effective accuracy of retrieval depends on the applied set of medical image retrieval techniques; therefore, the validation of such techniques is transversal to the architecture and independent from this paper's contribution. Nevertheless, the analysis of a complete solution's performance of retrieval, as well as the introduction of new ways to exploit the available information in a medical imaging archive, will undoubtedly be addressed in future work.

Appendix: Multimodal Query Schema

```
QueryNode {
  cokey: number (uint)
}
QueryFusion extends QueryNode {
  fusionOp: string | {
      name: string,
      ...options
    },
  children: array[QueryNode]
}
UnimodalQuery extends QueryNode {
  co: object {
    kind: string,
    queryText: string default "",
  }
}
TextQuery extends UnimodalQuery {
  co: object {
    kind: string = "text",
    queryText: string,
    keyword: optional boolean
  }
}
MediaQuery extends UnimodalQuery {
  co: object {
    kind: string,
    meta: array[object {
      key: string,
      value: string
    }]
  }
}
IndexedMediaQuery extends MediaQuery {
  co: object {
    uri: string
  }
}
StashedMediaQuery extends MediaQuery {
  co: object {
    uid: string
  }
}
```

References

1. Myers B: U.S. medical imaging informatics industry reconnects with growth in the enterprise image archiving market. 2012. [Online]. Available: http://www.frost.com/prod/servlet/press-release.pag?docid=268728701. [Accessed: 08-Feb-2016]

2. National Electrical Manufacturers Association (NEMA): Digital Imaging and Communications in Medicine (DICOM) standard. Rosslyn, VA, USA
3. Valente F, Viana-Ferreira C, Costa C, Oliveira JL: A RESTful image gateway for multiple medical image repositories. IEEE Trans Inf Technol Biomed 16(3):356–364, 2012
4. Akgül CB, Rubin DL, Napel S, Beaulieu CF, Greenspan H, Acar B: Content-based image retrieval in radiology: current status and future directions. J Digit Imaging 24(2):208–222, 2011
5. Müller H, Despont C: Health care professionals' image use and search behaviour. Proc Med Inform Eur. pp 24–32, 2006
6. Hanjalic A, Lienhart R, Ma W-Y, Smith JR: The holy grail of multimedia information retrieval: So close or yet so far away? Proc IEEE 4(96):541–547, 2008
7. Atrey PK, Hossain MA, El Saddik A, Kankanhalli MS: Multimodal fusion for multimedia analysis: a survey. Multimedia Systems 16(6):345–379, 2010
8. Valente F, Costa C, Silva A: Dicoogle, a PACS featuring profiled content based image retrieval. PLoS One 8(5):e61888, 2013
9. Müller H, Geissbuhler A: Medical multimedia retrieval 2.0. Yearb Med Inform 47(1):55–63, 2008
10. Cao Y, Steffey S, Jianbiao H, Xiao D, Tao C, Chen P, Müller H: Medical image retrieval: a multimodal approach. Cancer Inform, 2015
11. Mourão A, Flávio M: NovaMedSearch: A multimodal search engine for medical case-based retrieval. In Proceedings of the 10th Conference on Open Research Areas in Information Retrieval. Le Centre de Hautes Etudes Internationales D'Informatique Documentaire, 2013, pp 223–224
12. Hanbury A, Boyer C, Gschwandtner M, Müller H: KHRESMOI: towards a multi-lingual search and access system for biomedical information. Med-e-Tel, Luxembourg, 2011, pp 412–416
13. Schaer R, Markonis D, Müller H: Architecture and applications of the parallel distributed image search engine (ParaDISE). FoRESEE, Stuttgart, 2014
14. Widmer A, Schaer R, Markonis D, Müller H: Gesture interaction for content–based medical image retrieval. In Proceedings of international conference on multimedia retrieval, 2014, p 503
15. Markonis D, Donner R, Holzer M, Schlegl T, Dungs S, Kriewel S, Langs G, Müller H: A visual information retrieval system for radiology reports and the medical literature. In Multimedia modeling conference, 2014
16. Rahman MM, You D, Simpson MS, Antani SK, Demner-Fushman D, Thoma GR: Multimodal biomedical image retrieval using hierarchical classification and modality fusion. Int J Multimed Inf Retr 2(3):159–173, 2013
17. Valente F, Silva LB, Godinho TM, Costa C: Anatomy of an extensible open source PACS. J Digit Imaging 29(3):284–296, 2016
18. Costa C, Freitas F, Pereira M, Silva A, Oliveira JL: Indexing and retrieving DICOM data in disperse and unstructured archives. Int J Comput Assist Radiol Surg 4(1):71–77, 2009
19. Datta R, Joshi D, Li J, Wang JZ: Image retrieval: ideas, influences, and trends of the new age. ACM Comput Surv (CSUR) 40(2):5, 2008
20. Mourão A, Martins F, Magalhães J: Multimodal medical information retrieval with unsupervised rank fusion. Comput Med Imaging Graph 39:35–45, 2015
21. Fox EA, Shaw JA: Combination of multiple searches. NIST SPECIAL PUBLICATION SP, p 243, 1994
22. Cormack GV, Clarke CLA, Buettcher S: Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval, 2009, pp 758–759
23. MontagueM, Aslam JA: Relevance score normalization for metasearch. In Proceedings of the tenth international conference on information and knowledge management, 2001, pp 427–433
24. Lee JH: Analyses of multiple evidence combination. In ACM SIGIR forum, 1997, vol 31, pp 267–276
25. Müller W, Müller H, Marchand-Maillet S, Pun T, Squire DM, Pecenovic Z, Giess C, De Vries AP: MRML: an extensible communication protocol for interoperability and benchmarking of multimedia information retrieval systems. In Information technologies 2000, 2000, pp. 124–133
26. Markonis D, Holzer M, Baroz F, De Castaneda RLR, Boyer C, Langs G, Müller H: User-oriented evaluation of a medical image retrieval system for radiologists. Int J Med Inform, 2015
27. Rocchio JJ: Relevance feedback in information retrieval. In The SMART retrieval system, experiments in automatic document processing, 1971, pp 313–323
28. Markonis D, Schaer R, Müller H: Evaluating multimodal relevance feedback techniques for medical image retrieval. Inf Retr J, pp 1–13, 2016
29. Faruque J, Beaulieu CF, Rosenberg J, Rubin DL, Yao D, Napel S: Content-based image retrieval in radiology: analysis of variability in human perception of similarity. J Med Imaging 2(2):25501, 2015