Computers & Graphics (2022)

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Exploring scientific literature by textual and image content using DRIFT

Ximena Pocco[a,1], Tiago da Silva[b,2], Jorge Poco[b,2], Luis Gustavo Nonato[c,3], Erick Gomez-Nieto[a,1]

[a] *Quinta Vivanco s/n Urb. Campiña Paisajista, Arequipa, Arequipa, Peru*
[b] *Praia de Botafogo, 190, Rio de Janeiro - RJ, 22250-900, Brazil.*
[c] *Trabalhador São-carlense Avenue, 400, São Carlos-SP, 13566-590, Brazil*

ARTICLE INFO

Article history:
Received March 29, 2022

Scientific Literature, Search interfaces, Multimodal Processing, Visual Analytics

ABSTRACT

Digital libraries represent the most valuable resource for storing, querying, and retrieving scientific literature. Traditionally, the reader/analyst aims to compose a set of articles based on keywords, according to his/her preferences, and manually inspect the resulting list of documents. Except for the articles which share citations or common keywords, the results retrieved will be limited to those which fulfill a syntactic match. Besides, if instead of having an article as a reference, the user has an image, the process of finding and exploring articles with similar content becomes infeasible. This paper proposes a visual analytic methodology for exploring and analyzing scientific document collections that consider both textual and image content. The proposed technique relies on combining multiple Content-Based Image Retrieval (CBIR) components and multidimensional projection to map the documents to a visual space based on their similarity, thus enabling an interactive exploration. Moreover, we extend its analytical capabilities with visual resources to display complementary information on selected documents that uncover hidden patterns and semantic relations. We evidence the effectiveness of our methodology through three case studies and a user evaluation, which attest to its usefulness during the process of scientific collections exploration.

1. Introduction

One essential task of scientific research is the literature review. It seeks to identify, evaluate and synthesize published information in a specific subject or topic. Typically, it is performed by querying different academic sources, *e.g.*, journal papers, surveys, reviews, books, and theses/dissertations stored in digital libraries. For instance, well-known repositories such as IEEE Xplore[4], ACM DL[5], and ArXiv[6] enable the traditional searching paradigm where users perform queries based on keywords, resulting in a list of textual snippets containing the title, authors, and other information summarizing the content of each document. Users must manually inspect the snippets to find documents of interest; digital libraries do not provide resources to gather documents based on their content, making the literature compilation a tedious and time-consuming task. Moreover, resources to perform queries from images, tables, and charts are not available, impairing the search for content other than text. The image-based query has been widely used in Content-based

e-mail: `ximena.pocco@ucsp.edu.pe` (Ximena Pocco), `b41308@fgv.edu.br` (Tiago da Silva), `jorge.poco@fgv.br` (Jorge Poco), `gnonato@icmc.usp.br` (Luis Gustavo Nonato), `emgomez@ucsp.pe` (Erick Gomez-Nieto)

[1] Deparment of Computer Science. Universidad Catolica San Pablo, Arequipa, Peru

[2] School of Applied Mathematics. Getulio Vargas Foundation, Rio de Janeiro, Brazil.

[3] Institute of Mathematics and Computer Sciences. University of Sao Paulo, Brazil

[4] `http://ieeexplore.ieee.org/`

[5] `http://dl.acm.org/`

[6] `http://arxiv.org/`

Image Retrieval (CBIR) systems and could also be employed to support the exploration of scientific literature libraries. Performing queries based on images and other non-textual content can make it possible to answer questions such as: "Which are the typical images in papers from this author?", "Which articles have images similar to this one?" or even "Is this image similar to any other published?".

Another critical issue in exploring scientific literature is how to enable visual resources that render the analysis of multiple document collections an easier task. Some academic search engines such as Microsoft Academic Visual Explorer[7] and Google Scholar[8] enable visual representations for co-authorship analysis and citation evolution over time. Besides being quite limited, the visual resources enabled in those tools are not linked to query mechanisms, which considerably restricts the scope of any exploratory analysis. There are also alternatives to replace the regular list of textual snippets with some visualization-oriented representations, mainly in the context of web search result analysis [1, 2, 3]. However, despite the effectiveness demonstrated by these methods, they have not been introduced into digital libraries for exploring articles yet.

In this work, we propose an interactive visualization tool for exploring extensive collections of scientific documents. Called *DRIFT* (Document exploration based on Image and textual Features), the proposed methodology combines core CBIR functionalities with an interactive multidimensional projection mechanism that identifies documents with similar content, including images. In contrast to existing systems, our approach enables several exploratory and visualization resources that make complex analysis doable, increasing the user's ability to perform complex searches and analyses.

In summary, the main contributions of this work are:

- A methodology that combines content-based image retrieval mechanisms, multidimensional projection, and visual analytic tools into a single framework that handles documents based on their textual and image content.
- A visual analytic tool called DRIFT, which implements the proposed framework to enable customized exploration of collections of scientific documents.
- Three case studies and a user evaluation that demonstrate the utility and effectiveness of our methodology.

In a previous version of this paper [4], we showed the rationale behind DRIFT and how its components support the analyst in the exploration process. In this version, we extend this discussion by describing our case studies deeply, introducing a new case on Coronavirus (COVID-19) research, testing a new strategy for textual processing, and detailing the feedback provided by users after the evaluation process.

[7]`http://academic.research.microsoft.com/VisualExplorer`
[8]`http://scholar.google.com/`

2. Related Work

We focus the following discussion on methods that explore scientific publication collections. Therefore, we group existing methods into three main categories, *i.e.*, citation-based, textual-content-based, and image-content-based. We briefly describe some relevant techniques from the first group, however, we focus this section on both textual and image-based methods since they are the foundation for our work. A more comprehensive review of visualization methods to explore scientific document collections can be found in [5].

Citation based methods focus on uncover citation and research collaboration patterns [6, 7, 8, 9, 10]. Liu *et al.* [11], for instance, search for citations in a specific paper and build a tag cloud to intuitively convey which part of the paper each citation refers to. Yan and Ding [12] analyze six different types of scholar networks (coupling, (co-)citation, topic, co-authorship, co-word) aiming a better understanding of how they are related. PaperPoles [13] extract references/citations from some seed articles for ordering them by relevance to positive or negative queries. cite2vec [14] makes use of word embeddings for document exploration based on the context in which they are cited. Recently, doccite2vec [15] proposes a model for paper recommendation by gathering citations and document embeddings. Although those methods allow an informative analysis of authorship relations, the information extracted is not plenty to characterize publications in order to generate insights.

Textual content-based methods make use of text processing strategies to establish similarities among documents. For instance, Action Science Explorer (ASE) [16] is a system that enables interactive analysis of a paper collection through linked-views, identifying key papers, topics, and research groups. The integration between text analysis and citation context turns ASE into an informative representation. However, the visualization suffers from the problem of occlusion, as textual labels can overlap. Survis [17] is a visual analytic system designed to analyze and disseminate literature databases. A set of linked views allows users to explore citation relations over time. One remarkable feature is the use of an interactive selector for enriching visualizations, providing a visual mechanism for ordering and filtering publications. MIST [18] employs keywords from scientific documents to generate semantically aware and overlap-free word clouds. PEx-Web [19] is an interactive tool that relies on multidimensional projections to map web search results, including patent collections, by similarity into a 2D point-based layout. VisIRR [20] uses a 2D scatterplot to visualize and recommend documents based on user preferences. Literature Explorer [21] uses standard visual components such as trees and theme river to detect thematic topics to support document retrieval, avoiding that the number of topics has to be pre-defined.

Image-based methods comprise a class of methods that aim to extract and process images from scientific documents, which are then employed to query and compare scientific documents. One of the few image-based approaches described in the literature is the work by Deserno *et al.* [22], which makes use of images with annotated words to query and group medical documents, reporting a gain in the quality of the query due to the use of images. In fact, the benefit of using images to enrich the querying process has also been reported by Muller *et al.* [23], showing that the relevance of documents retrieved from text and images is higher than using only textual queries. Commercial tools also are part of this group, as in the case of Pinterest [24], eBay [25], and Alibaba [26], which faces the big challenge of searching into image large collections by using deep neural network models.

In a lower number, some approaches are devoted to combining two or more methods. For instance, Felizardo et al. in [27] and [28] uses graphs and edge bundles to understand how a network of articles references each other in the collection while examining the textual content of the articles by a multidimensional projection method. However, it falls in a visual occlusion when it scales in a number of documents, especially in its citation map, impairing the exploration of large datasets. Papercube [29] is a web-based application that integrates timelines, treemaps, and graphs to represent article citations and metadata. In the same context, PaperVis [30] presents a mixed representation based on keywords and citations for exploring scientific papers. One of its most valuable contributions is the introduction of a tree-based mechanism to visually review the visualization history. Despite that, none of them allow us to manipulate these interactions from the user activity to generate new insights and support the exploration task.

The method proposed in this work combines the last two approaches discussed above, enabling interactive linked components to efficiently uncover hidden relation patterns in scientific document collections. Moreover, DRIFT allows analysts to restore and compare previous states of his/her interaction, helping the construction of insights from different selections. DRIFT turns out to be useful in several tasks, as the quick identification of papers of interest and analysis of their content.

3. Goals and Analytical Tasks

To define our goals, we had a series of meetings with multiple researchers with 5 to 15 years of experience. All of them are professionals in different fields of study but into STEM disciplines. All meetings consisted of individual interviews focused on the advantages and limitations of the current paradigm used by digital repositories and the participant's experience using these. Also, we conducted an exhaustive literature review to evaluate available systems for scientific literature exploration. As a result of this, we came up with a set of goals and analytical tasks that guided our tool design.

3.1. Goals

Below we describe the four objectives that lead to the development of our tool.

- **G1. Support exploration of scientific documents collections.** Available digital libraries offer limited tools for analyzing scientific documents since their exploration relies on the accuracy of its search engine for retrieving relevant documents. However, researchers could not have exact inputs for performing accurate queries, requiring an exploratory analysis to know about the collection and extract significant insights. Our goal is to build a visual analytic tool that enables scientific document collection exploration by combining a set of interactive resources and allowing the analyst to identify documents of interest. In this way, digital libraries might benefit from this proposal.

- **G2. Integrate image and textual content.** Most scientific literature exploration tools focus on text to organize documents — *e.g.*, text matching, citation networks — preventing the exploitation of several features available in scientific papers. Thus, one main goal for our project is to build a tool to perform a multimodal exploration. For that, we wish the analyst to query for both image and textual content to lead the exploration process.

- **G3. Understand metadata and topics in documents groups.** Researchers are quite familiar with reviewing document metadata — *e.g.*, authors, publisher, and publication date — since it provides additional information to decide about document relevance. Likewise, recognizing topics rapidly from document groups enhances analysts' capabilities to review more literature. We identify this goal as an opportunity for improving the manner how researchers can effectively extract insights from document collections while interactively refining their search criteria.

- **G4. Support literature review task.** Exploring and analyzing scientific literature end up in customized collections containing relevant documents for the analyst. These collections organize references according to specific interests and motivations. For instance, support the writing of the Related Work section for an article or prepare a bibliography for a curricular syllabus.

3.2. Analytical Tasks

After understanding the goals of the project, we define the set of analytical tasks that our tool must support.

- **T1. Image similarity queries.** Given a query image, we want our tool to be able to retrieve a set of images ranked by similarity. These results allow the analyst to discover documents associated with the retrieved images. This task supports goals **G1** and **G2**.

- **T2. Group documents based on image and textual content.** Enable analysts to group documents and create collections considering both textual and image content. This task allow us to achieve goals **G1** and **G2**.

- **T3. Selecting and filtering collections.** Allow the analyst to select a document collection and filter its content (*i.e.*, adding or removing documents) according to his/her search criteria. This task allow us to achieve goals **G1** and **G2**.

- **T4. Compare document collections.** Enable topics and metadata analysis in document collections created by the

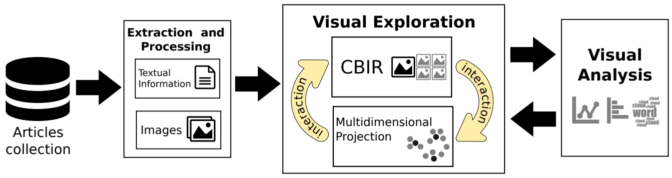


Fig. 1: The proposed methodology comprises three main steps: Extraction and processing of image and textual information from the documents, content-based image retrieval (CBIR) and interactive multidimensional projection (middle), and visual analysis to uncover hidden patterns and relationships among subsets of documents.

analysts. Moreover, our tool must facilitate comparison between document collections to identify similarities among them. This task gives support to goal **G3**.

- **T5. Storing and managing document collections.** Each time an analyst creates a document collection, he/she must be able to save it. Moreover, users should have access to the stored collections to perform operations such as querying, retrieving, and merging. This task allow us to achieve goals **G3** and **G4**.
- **T6. Exporting customized document collections.** After the exploration process, the analyst should be able to export his/her results into a human and machine-readable format. This task allows us to achieve goal **G4**.

4. DRIFT

DRIFT is a visualization tool designed to support the analysis and exploration of large scientific document collections, revealing the similarity between document contents while enabling interactive resources to store and recover intermediate steps of the exploratory analysis. DRIFT's methodology, illustrated in Fig. 1, comprises three main steps: (i) extraction and processing of image and textual information from each document, (ii) interactive exploration of a multi-CBIR, and (iii) visual in-detail analysis of selected documents.

DRIFT allows the analysts to choose the number of images to be used for querying as well as the number of images to be retrieved by the CBIR components. Each component brings a set of documents associated with the images, *i.e.*, the documents that contain the retrieved images as part of their content. The associated documents are considered as control points to guide the multidimensional projection process, which is responsible for mapping the documents based on their similarity to 2D visual space. Textual features are only used to accomplish the projection. Thus, using images of interest users can find relevant documents that are then used to drive exploratory analysis in a 2D visual space. Additionally, we implement visual resources to support analytical tasks, *i.e.*, author-frequency, and year-frequency histograms as well as a topic-based word cloud. A streamgraph component, called Selection Visual Manager, helps to save and display intermediate steps of the exploratory analysis. Such intermediate steps, called *states*, can be recovered, compared, and employed to generate new states, which can be downloaded as subsets of documents.

Table 1: Methodological and analytical properties and their related tools.

	T1	*T2*	*T3*	*T4*	*T5*	*T6*
Multi-CBIR View	•					
Multidimensional Projection View		•	•			
List-based Selection Refinement View		•	•	•		
Selection Content Summarization View			•	•		
Selection Inspector View					•	•
State Manager View					•	•

We design these visual components to achieve the identified goals. All components address at least one analytical tasks described in Section 3.2. Table 1 details the relation between the visual resources and analytical tasks (T1-T6 columns).

4.1. Content Extraction and Processing

We make use of ArXiv® digital library and the e-proceedings of a well-known conference in visual computing from 2011 to 2014 as the document collections to be handled by DRIFT. We divided the collection into three subsets, namely, (DT1) containing 1369 articles on five distinct topics, (DT2) containing 171 visual computing articles, and (DT3) containing 284 articles on three related topics. The main reason for this division is to provide different scenarios that will make it easier to assess the performance of our methodology. The keywords used as well as the number of images contained in each category is described in Table 2.

The tool *pdf2text* from Poppler library[9] is applied to convert the textual content of PDF files into ASCII text files. ASCII files are then analyzed as described in Section 4.1.1. The tool *pdfimage* also from the Poppler library is used to convert pages of PDF documents into 8-bit PNG image files. The PNG files are input into an image processing pipeline to extract figures contained on each page. We use the global Otsu method [31] to binarize the PNG files, searching for components in the resulting binary images. The components are then ranked according to their area (*height*×*width*) and the largest component in an image is considered a figure if its dimensions are greater than 50×150 pixels. In this case, the corresponding bounding box is cropped out from the original PNG file and exported as an individual image to a local image database. The detected figure is then erased from the image page and the whole process is repeated until no component satisfying the size criterion is found. Then we move to the next PNG file. Finally, the saved figures are submitted to feature extraction as described in Section 4.1.2.

4.1.1. Text Processing

We employed two approaches to perform suitable text processing. First, as proposed in the previous version of this

[9] `http://poppler.freedesktop.org`

work [4], we adopt the text processing procedure proposed by Gomez-Nieto *et al.* [32] to extract textual features. The choice is mainly due to simplicity and good computational performance in processing short-length text. In summary, ASCII files associated with each document are processed to extract term frequency vectors. Then, we performed some conventional text processing filtering — *i.e.*, stemming, stop word removal, and definition of Luhn's lower and upper cuts [33] — ending with a TF-IDF vector representation of each document.

Second, we introduce embeddings to extract each document's representative vector in this extended version. Specifically, we choose doc2vec [34], an unsupervised model that seeks to understand each word's context in a document and find similarities between documents. Thus, we instantiated the model using Gensim [35] library, with a minimum word count of 2, a vector size of 50, and the number of training iterations of 40.

In both cases, we process only the abstract rather than the full content of each document to reduce the computational burden. Although the entire document content could improve the quality of the document representation, handling only the abstracts favors interactivity. It makes it easier to plug the proposed solution into a web environment.

Note that using one of these two methods aims to provide suitable input for our multidimensional projection step (presented in Section 4.2). However, DRIFT proposes a methodology independent of a unique method to perform this task, and any other adequate method can be employed.

4.1.2. Image Features Extraction

Over the last 5 years, deep-learning-based feature extraction techniques have been unbeatable in many different contexts, as for example in image object detection. Therefore, we use a neural network to extract feature vectors for each image extracted from the documents. We relied on AlexNet [36] architecture pre-trained on the ImageNet dataset[10]. Unlike the original network architecture, we changed the last layer from 1000 to 5 neurons. It was done to fine-tune on our first dataset (DT1) which consisted of 5 categories (disease, gene, gravitational, market, seismic). We then modified the learning rate of the last layer by a factor of 10; this allows the back-propagation to have a high effect on the last layer and a slight impact on the previous ones. Finally, we did 50,000 iterations with a momentum of 0.9 and a base learning rate of 0.001. The features extracted from the last fully connected layer are a vector of 4,096 elements.

For the other two datasets (DT2 and DT3) we did not fine-tune the CNN because the images did not contain classes, that is why we extract the feature of these two datasets using the model already trained with DT1. Note that although we are doing fine-tuning in a classifier, our goal is not to use the classifier output, but to refine and extract the characteristics for our problem. To avoid the curse of dimensionality, we additionally reduce the feature vector dimension to 50 using PCA.

[10] `http://image-net.org/`

Table 2: Description of datasets used for our study: Query used, number of documents retrieved (docs), number of images contained (imgs), textual processing strategy used (textproc), and source.

ID	Query	#docs	#imgs	textproc	Source
DT1	*seismic*	274	5,002	TF-IDF	ArXiv
	market	273	2,772		
	gravitational	274	3,082		
	disease	274	3,795		
	gene	274	3,731		
DT2	Proceeding 2011	45	1,010	TF-IDF	IEEE Xplore
	Proceeding 2012	45	1,195		
	Proceeding 2013	36	869		
	Proceeding 2014	45	1,033		
DT3	COVID-19	802	5392	doc2vec	Semantic Scholar
DT4	computer graphics	95	2,010	doc2vec	ArXiv
	image processing	93	1,922		
	computer vision	96	2,732		

4.2. Multi-CBIR and 2D Mapping

In the following, we describe the two main views that lead the exploration process and how they are integrated for interactively finding key documents.

Multi-CBIR View

Traditionally, a CBIR mechanism returns a similarity-based ordered list of images by querying one specific image. The similarity can be defined from a distance measure between feature vectors. In our implementation, the CBIR retrieves a user-defined number of similar images, which are displayed next to the query image in an imageboard, as illustrated in Fig. 2 (left). Up to five queries can be performed simultaneously using different input images. On the imageboard, images belonging to the same document are highlighted when the user hovers a specific image, fading out the remaining images on the board. The imageboard is the starting point for an interactive exploration of a document collection (see Fig. 3a).

Multidimensional Projection View

Textual information from each document gives rise to a high-dimensional vector that represents the document. In order to interactively explore the documents based on their similarity, we map the high-dimensional vectors to a 2D visual space using a multidimensional projection method. Specifically, we use Local Affine Multidimensional Projection (LAMP) [37] due to its interactive capability and good performance in terms of accuracy [38]. This method uses a reduced number of sample points (called control points) to drive the mapping of the remaining data instances into the visual space. LAMP makes it possible to interactively position control points on the visual space, updating the projection layout according to user intervention. This main feature is decisive for our choice of using LAMP over any other local multidimensional projection method as t-SNE [39] and UMAP [40]. We located our multidimensional projection component in the middle of the visualization interface, as shown in Fig. 3b.

Finding Key Documents

DRIFT combines multi-CBIR and multidimensional projection views to enable an interactive exploration of document collections. As illustrated in Fig. 2 a user query for images using the CBIR component and the result are presented in the imageboard. Documents associated with retrieved images are set up as control points to drive the multidimensional projection step. Once projected, control points can be dragged and dropped to emphasize their semantic relation, being the projection layout updated accordingly. A parameter α can be tuned between zero and one $(0-1)$ forcing LAMP to perform a more local or global mapping, respectively. Images resulting from the CBIR and their associated documents are highlighted in the projection layout using the same colors of the groups in the imageboard.

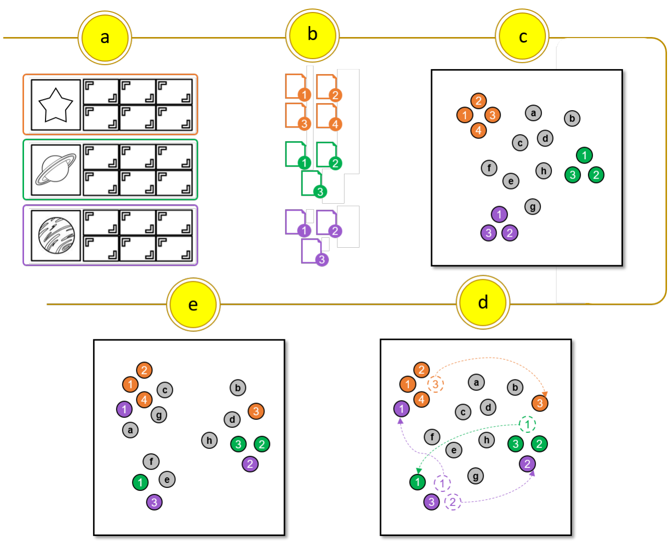


Fig. 2: Finding key documents by the combination of multiple-CBIR and multidimensional projection views: (a) Our system retrieves multiple sets of images based on image queries, (b) automatically it selects the documents where retrieved images are contained, (c) these documents are used as control points by our multidimensional projection view for mapping the entire document collection, (d) according to his/her interests the analyst repositions the control points to customize groups, and finally (e) the entire collection is reprojected based on such reposition.

4.3. Visual Analysis

The main functionality of our methodology is the interactive selection of subsets of scientific documents. The user can select a subset of articles by drawing a polygon around points (documents) of the projection. The borders of the points selected will be colored in red. Each time the analyst selects a subset of documents, linked views are updated showing relevant information from the selected documents. Relevant information is depicted in the following visual components:

List-based Selection Refinement View

This component shows the list of selected documents, depicting the title and DOI, where the latter is linked to the original publication page. Particular documents can be removed from the list by clicking on the trash button, as shown in Fig. 3c.

Selection Content Summarization View

Once a group of documents is selected, three visual summarization widgets are updated to show relevant content from the selected subset. Specifically, the visual summarization widgets show author-frequency histogram, topic word cloud, and publication-year frequency histogram, as shown in Fig. 3d. Those widgets provide an overall view of most-cited authors, general/particular topics discussed, or which period comprises the larger number of publications.

Selection Inspector View

A typical cycle accomplished several times is selecting documents and inspecting their summary to choose the most relevant ones. However, along with the exploration, some of these selections can partially reveal relevant documents for the user. To take advantage of this new subset of documents, DRIFT allows the user to manage such selections by performing recover, compare, and merge operations. We build a visualization-assisted mechanism that organizes and saves each selection from Multidimensional Projection View as a state to facilitate such an iterative process.

We employ an interactive streamgraph metaphor that stores the documents, images, and authors resulting from each iteration cycle. The number of documents, images, and authors are represented as streamgraph layers — orange, red, and turquoise, respectively — and each iteration cycle is marked with three vertically aligned dots in the layout, as illustrated in Fig. 3e.

This widget allows recovering a state saved during the analytical process, supporting to restore relevant articles identified during any cycle. Indeed, the widget enables a wide range of operations over, *e.g.*, compare, combine, or delete the result of any iteration cycle, as detailed in the following view.

Our choice for employing this metaphor was motivated by the following requirements: (i) explore temporal information that can be drastically scaled by the number of user selections, (ii) rapidly inspect how the number of documents/authors/images varies about previous states, and (iii) an overlap-free representation that allows us to analyze by attribute and state simultaneously. These needs justify our choice over traditional charts (e.g., line charts, bar charts, or boxplots) that impairs readability as this visual resource scales in terms of the amount of data and area occupied.

State Manager View

Suppose that during the exploratory analysis two states (S_A and S_B) are produced. To compare the content of the two states DRIFT employs a modal window that performs set operations on states S_A and S_B: intersection $(A \cap B)$ and difference $(A - B$ or $B - A)$. After selecting the two states from the streamgraph and clicking on the "compare" button the modal window shows up, as illustrated in Fig. 4. The modal window is divided into three horizontal blocks, one for each set operation. Each block contains the title, authors, and images of each document resulting from the set operation.

The result of a set operation can be saved as a new state in the streamgraph. On the bottom part of the modal window, under

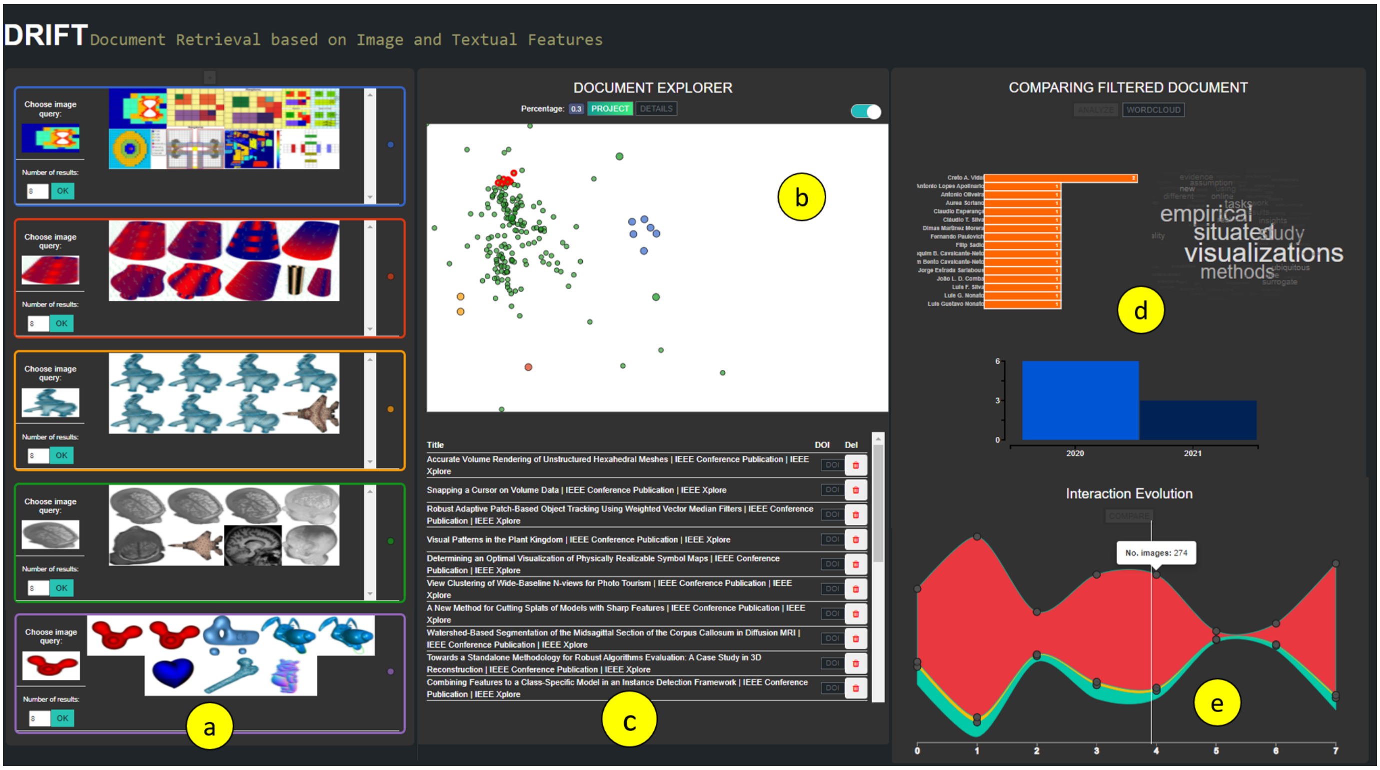


Fig. 3: An overview of DRIFT tool. In (a) the imageboard resulting from multi-CBIR queries. The output of this component is a set of images related to the query images. The second column is divided into two sections: (b) the projection of the scientific documents based on their similarity. Control points driving the projection are represented by circles with a larger diameter using the same color as the imageboard to emphasize the correspondence between images and documents. Users can select a set of documents of interest and see in detail their content in the lower view (c). In (d) we show a histogram of the authors, a word cloud, and the publication year histogram. In (e) the selection made by the user is presented into the Selection Inspector View.

title *Selected*, the title of chosen documents is displayed. Once the *New state* button is selected, a new state will be added to the streamgraph. The user can also export the filtered documents as *.json* files containing the selected article titles and their respective web links.

Our prototype is entirely developed in Javascript, using D3.js[11] and Lasso[12] libraries, what should make it possible to plug DRIFT into digital libraries running on Web. The preprocessing steps such as feature extraction are speeded up using C++ standard libraries.

5. Case Studies

In this section, we present three case studies to assess DRIFT's effectiveness in terms of exploration of the scientific document collection. Each one of them represents a different scenario. The first involves the exploratory analysis of the dataset DT1 (see Table 2) where queries are performed from five different topics, namely, *"seismic"*,*"market"*, *"gravitational"*, *"disease"* and *"gene"* (the ArXiv digital library is the collection). The second analysis involves the dataset DT2 (Table 2) where documents proceeds of the *Conference on Graphics, Patterns and Images (SIBGRAPI)* from four specific years, namely, 2011, 2012, 2013, and 2014. Finally, the third involves the dataset DT3 (Table 2) which contains a collection of articles related to research on Coronavirus (COVID-19).

5.1. Exploring the DT1 Dataset

Suppose we are looking for articles related to gravitational waves associated with supernovae. We start the exploratory analysis by performing queries from images related to the topic of interest. Fig. 5 shows two images used as input for the querying process. To emulate the behavior of an analyst in this topic, we select these two inputs knowing *a priori* that they appeared in articles that talk about the topic searched. The first input image (top) is related to a novel gravitational-wave signature in supernovae and we decide to retrieve six images related to the given one. The number of retrieved images is a user-defined parameter, and the choice for six was made due from the seventh image onwards they do not belong within the domain we are looking for. We found three articles related to seismic features and gravitational waves, as shown in the list of documents associated with the retrieved images. Using the second input image (Fig. 5 bottom) and setting the number of retrieved images to nine the search results in five documents.

The eight documents are used as control points to drive the mapping of the entire document collection. Fig. 6 illustrates the entire interactive analytical process. Each disk encloses a state saved and represented in the streamgraph, showing the projection point cloud, selected documents associated with the state,

[11] http://d3js.org/
[12] http://github.com/skokenes/D3-Lasso-Plugin

Fig. 4: Comparing two different states of user interaction by using the State Manager View: (a) shared articles between A and B selections, (b) articles present in selection A and not in B, (c) articles present in selection B and not in A, and (d) filtered articles for generating a new state or exporting to file.

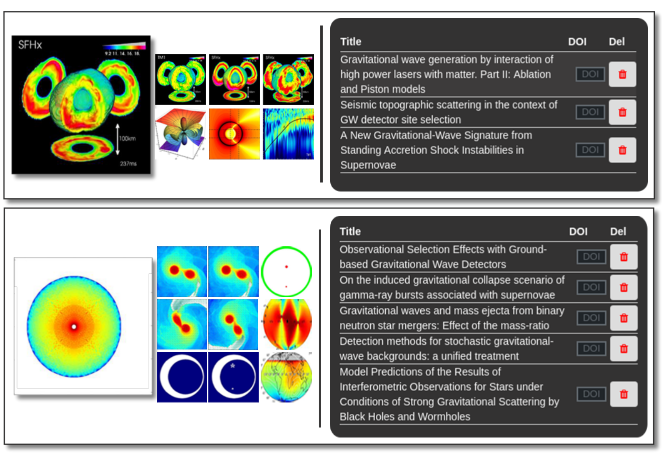


Fig. 5: Two input images to start the exploratory analysis of the DT1 dataset.

and the three summarization widgets. Initially, the projection in the first state displays two sets of control points, the blue point group on the right side from the first query, and the red point group on the left side from the second query. Notice that the word cloud summarization widget by selecting the blue points is basically formed by the words "gravitational", "frequency" and "scattering". In the same way, the second state is the selection of points of the left side. States 3 and 4 are composed by selections with non-relevant documents for our analysis. In-state 5, the inner region of the projected point cloud is selected, revealing a broader range of topics published from 2007 to 2016. Up to here, these selections allow identifying the topics around the different regions of the projection. At this point, we look to determine which control points we must relocate closely and which distantly. On the ninth state, the projection is locally modified (parameter $\alpha = 0.3$ is set to 0.001 in LAMP). Additionally, one control point is moved on the bottom-right (blue) and another on the bottom-left (red) to better separate documents deemed relevant from those of low interest. Documents close to the target control points reveal a large number of publications sharing co-authors, especially in 2008, 2014, and 2016 years. The analyst can determine which control point is approaching its target by hovering it with the cursor and watching the image(s) that are highlighted in the CBIR views.

States 6 to 9 comprise different document selections on the same projection. As the word cloud generated in the ninth state reveals, it includes several topics. At this point, we are interested in comparing the current selection and selection stored in the previous state (state 8). For that, we make use of the Selection Inspector View, and after a quick look at each imageboard and document title, we filter out eight relevant articles, giving raise a new state in the streamgraph (state 10).

On the eleventh state, control points are moved even further, being placed on the top-left region where gravitational, mass, and accretion topics reside. On the twelfth state, documents on the rightmost region of the projection layout are associated with topics of interest. However, some documents clutter the analysis, so we resort to the managing states tool to compare the current and the ninth state. The resulting analysis gives rise to a subset of nine documents saved in the thirteenth state, which are mostly related to *"gene"*, *"data"*, and *"model"*. Finally, we decided to combine the two states deemed most relevant for our analysis, the eleventh and thirteenth states. We use one last time the managing state tool, resulting in a set of articles closely related to the topics of interest, namely *"gravitational"*, *"wave"*, *"simulations"*, *"scattering"*, which have been published in 2011, 2013, 2015, and 2016, this later with a larger number of publications. The merged states are export as a JSON file for future analysis and readings.

5.2. *Exploring the DT2 Dataset*

In the second case, we aim to find articles related to 3D models (see Fig. 7) starting with five query images. We chose four of them by their explicit relation to our target, and the last from another topic, *i.e.*, a well-known picture in the context of image

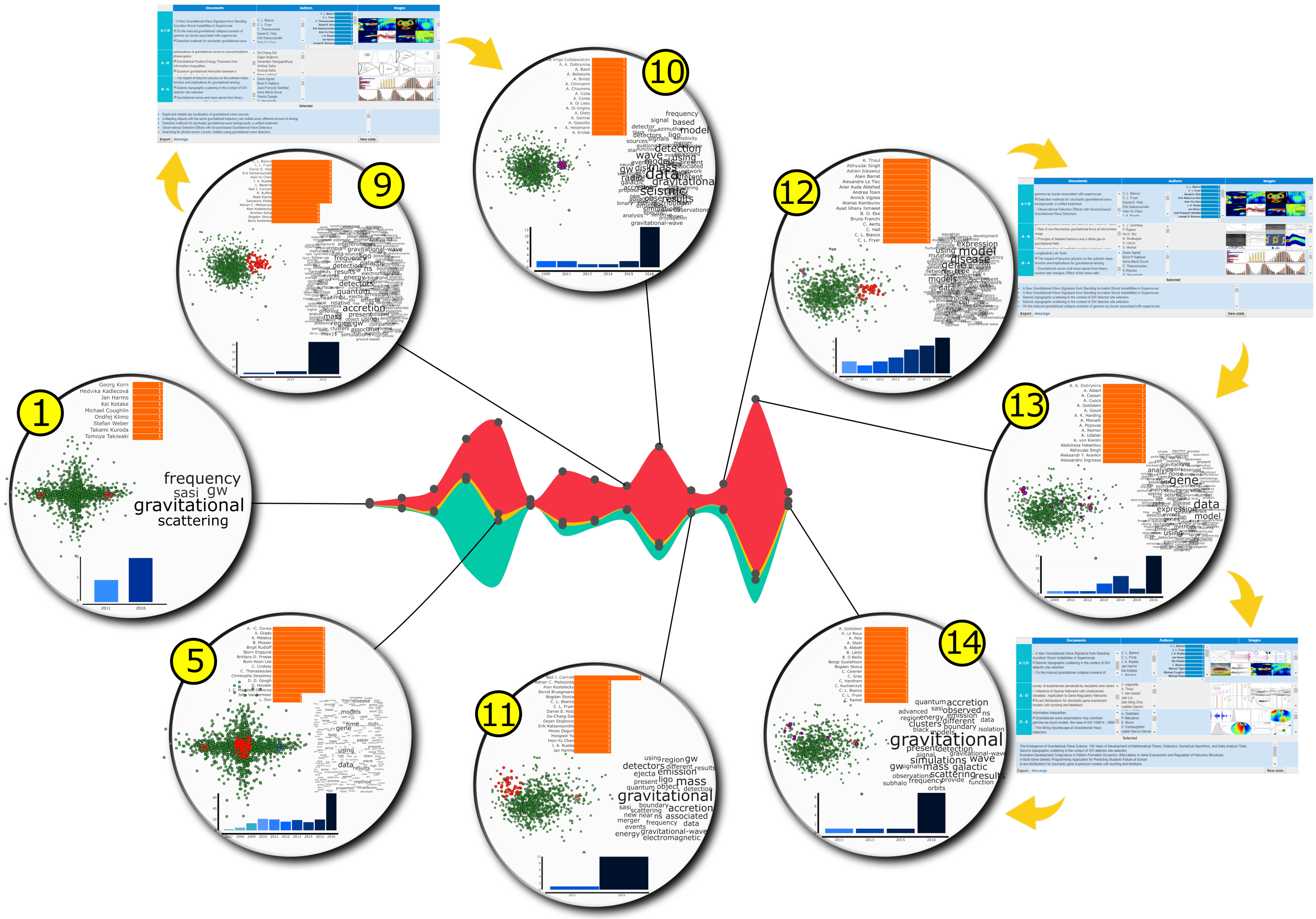


Fig. 6: Exploring 14 different states on the streamgraph widget during user exploration of DT1. Each colored layer represents the evolution of the number of (■) images, (■) documents, and (■) authors by each selection. Additionally, we highlight eight states for showing the set of selected documents on projection, two histograms containing top authors and publication year, and the resulting word cloud. The order number of these states is displayed in a yellow circle on the border.

processing. The main purpose of using such an image is to employ it as a control point to properly drive the multidimensional projection, pushing unrelated articles towards this control point.

The initial projection places most instances in the middle of the layout, as illustrated in Fig. 7a. Using the interactive selector — which displays the title of the article — one can easily see that documents placed at peripheral regions belong to distinct topics, as shown in Fig.s 7b and 7c, respectively.

We rearrange our projection by moving a few control points, *i.e.*, one blue control point to the bottom-left region, and the only red control point to the right region, as shown in Fig. 7d. Such an operation map some points around the recently reallocated blue control point. When we inspect for the content of these points (using the word cloud component) we notice important terms for our search, as *"face"*, *"reconstruction"*, *"3d"*, *"skull"*. In the same way, in Fig. 7e we gather red, blue, and violet control points, and select them and their neighbors. The generated word cloud display terms as *"gestures"*, *"deformation"*, *"mesh"*. Both selections reveal two different configurations, in Fig. 7d we group some documents that talk about 3D mostly, while Fig. 7e depicts images and words conveying image processing context.

Then, we aim to explore the content in some different regions of the projection, so after one more interaction, we found selected two groups for analysis. In Fig. 7f we highlight (orange and purple borders) these selections which contain partially related articles to our search. Both of them were stored in our *Selection Inspector View* as states 2 and 3 respectively, as illustrated in Fig. 7g. After interacting in our *State Manager View* we filtered a few articles to compose a new state, stored as state 4.

We relocate a few control points once again to refine our selections. In this step, we gather the orange, one blue, and one violet points on the middle-left region to separate a subset based on retrieved images from inputs, and leave on the right region control points already explored. As a result, in Fig. 7i we show the projection map with the neighbors selected around of such control points.

By simple inspection, we can notice that most of the articles retrieved depict similar textual content to the selected control points. We store this new selection as state 5. Then, we decide to analyze the contribution of one of the control points in orange which talks about curves on surfaces, so we drag it towards the middle region, bringing with it the most similar documents, as illustrated in Fig. 7j. As can be noticed, the neighbors generally talk about geometry processing for surfaces, which is close, but not completely, related to our search. We store this selection as state 6 in Selection Inspector. We opt to compare the first two selections (states "0" and "1") since they were not been carefully explored yet. In Fig. 7k, the *State Manager View* shows two selections without intersections. However, inspecting titles, authors, and contained images we filtered four useful articles for our purposes. We store the combination of these articles into state 7.

Finally, we compare states 5 and 7. We found two selections without intersection but containing four articles highly relevant to our study, as illustrated in Fig. 7l. At the end of our exploration, we have produced three states containing scientific articles that allow us to extract related methods to 3D modeling in computer graphics, *i.e.*, fourth, sixth, and eighth states. As can be noticed, we successfully discriminate such articles, even in a highly related-topic collection, by using images and textual information included in each article.

5.3. Exploring the DT3 Dataset

During the COVID-19 outbreak, different pieces of research were developed on the diagnosis of this disease. Among them, several proposals employed X-ray images to classify them into infected or non-infected patients. This case study aims to identify the documents that exploit this methodology for COVID-19 diagnosis, focusing on respiratory infections. We performed a search using the terms "coronavirus" and "infection" on the sentence-level search engine Spike[13]. As a result, we collected a corpus of 802 papers about COVID-19 infection published during 2019 and 2020, named DT3 in Table 2.

We started our study using two X-ray images of moderately compromised lungs, as Fig. 8a shows. Our choice is due since we knew *a priori* they proceed from a patient with COVID-19. The documents containing the CBIR outputs are mapped as red and blue points respectively, followed by the rest of the corpus in green, as illustrated in Fig. 8b. Initially, we select the red points to explore the content of such articles. Then, we perform a new selection containing the blue points and a few of their neighbors. Both are stored as states 0 and 1 in the Selection Inspector view.

After that, we filtered the states 0 and 1, labeled as A and B respectively, using the State Manager view and identified all articles related to respiratory affections. We can inspect these articles in Fig. 9a. At this point, we decide to exclude two documents from the row $B - A$ since they focus on transplants and lessons learned during the outbreak. As a result, we produce state 2 of our exploration. Interacting again with the projection, we selected articles near the red points, as shows in Fig. 9b. The resulting word cloud contains terms like lung, respiratory, and drugs. It provides a clue on respiratory infections, so we store this selection as state 3. Finally, we combine the states 2 and 3 intuiting that they will result into a more accurate selection. Thus, we obtain a state 4 composed by relevant 9 articles related to lung lesions, pneumonia, lung ultrasound, chest X-ray features of COVID-19, but sharing respiratory infections as a common topic, as shown in Fig. 9c.

As can be noticed, starting with just two images, DRIFT allowed us to identify nine other papers successfully. Additionally, all control points are associated with our search, so it was unnecessary to relocate them since closer regions already contain articles on respiratory infections. A text search would not necessarily have found articles with images similar to the one entered. With DRIFT, we exploit the characteristics of the X-ray image, obtaining valuable images from the beginning for a specialist in the area. Also, using DRIFT components to complement the exploration process, we reached a suitable selection of documents strongly correlated to our aimed subject – the identification of respiratory disease related to COVID-19.

6. User Evaluation

We conducted a controlled user evaluation to assess whether DRIFT enables the discovery of documents of interest in plausible time in comparison with the list-based traditional paradigm. In this section, we detail the procedure and results of our evaluation.

The evaluation follows a five-step procedure:

I *Introduction*. We gave a brief explanation of the purpose of the study to the participants.
II *Tool exposure*. We show the participants the necessary functionalities in DRIFT.
III *User familiarization*. Participants had 10 minutes to play with the tool, exploring a collection other than DT4.
IV *Evaluation*. We invited the participants to perform a specific search activity.
V *Feedback*. We asked to participants to bring us feedback from their experience using DRIFT.

We set-up two search activities (named A1 and A2) by using two specific questions, detailed in Table 3. Each activity involved an analytical procedure from the DT4 dataset, which contains 284 articles (with 6,664 images) extracted from ArXiv. These documents are from three fields of study: *Image Processing*, *Computer Graphics* and *Computer Vision*, as detailed in Table 2.

For this study, we invited six users, all experienced in systematic literature review and part of the initial group meetings participant for goals and analytical tasks definition — four with a master's degree and two with a doctoral degree — working on image processing, machine learning, or visualization from different research institutions. We split all participants into two groups containing three of them in each one, *i.e.*, group GR1 containing users 1 to 3 (U1-U3), and group GR2 users 4 to 6 (U4-U6), and where each group contains one participant with a doctoral degree. Then, we ask each group to perform the first search activity (A1) as follows: group GR1 using DRIFT, and group GR2 the list-based paradigm. Later, we ask to perform the second activity (A2) inversely, *i.e.*, group GR2 using DRIFT and group GR1 the list-based. We implement our list-based interface emulating the most traditional scientific repositories. Before our interface displayed all documents from DT4,

[13] https://spike.apps.allenai.org/datasets/cord19

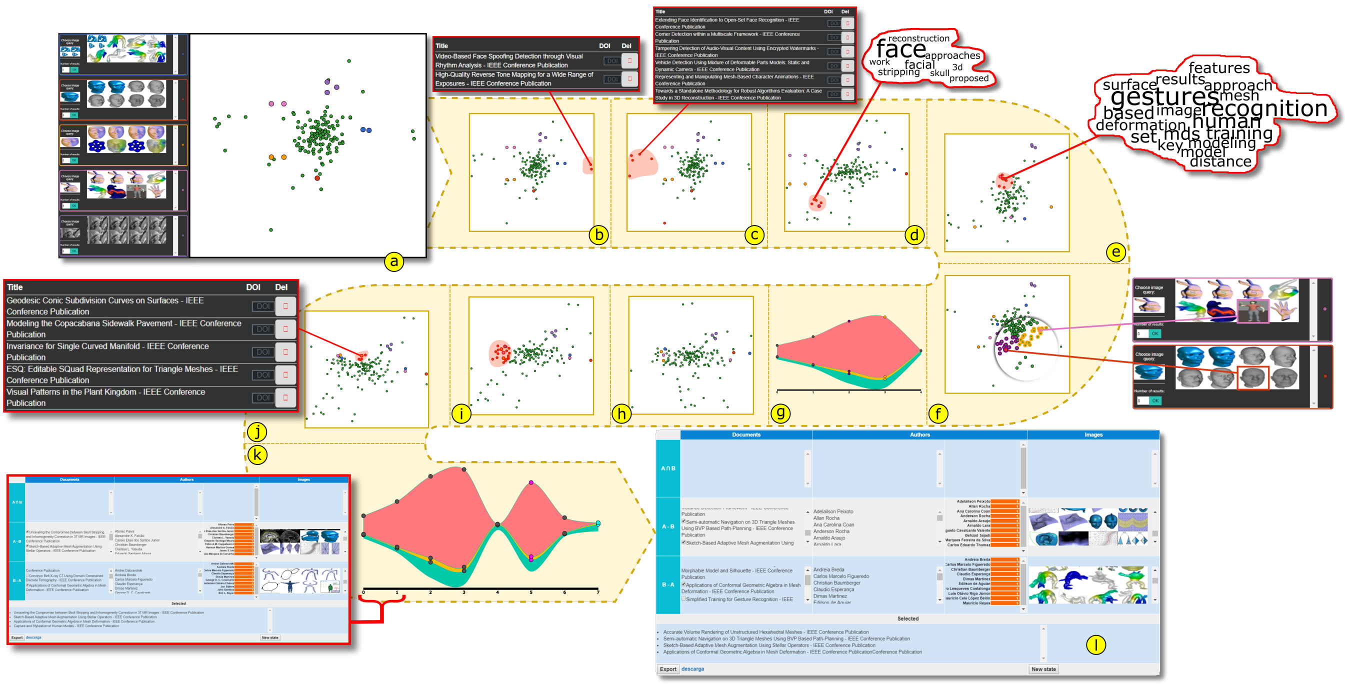
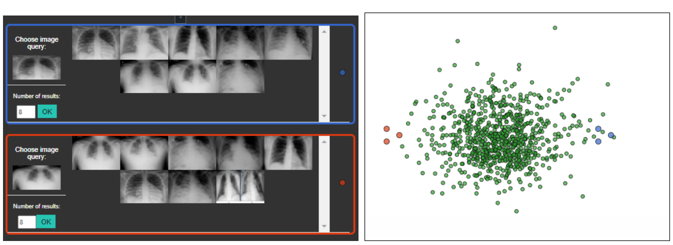


Fig. 7: Summarizing interactions in our case study on DT2: (a) input images and initial projection, (b-k) multiple interactions that include point reposition, comparison among states and inspection of visual resources, and (l) final selection of searched items.



(a) Inputs to multi-CBIR

(b) Multidimensional projection

Fig. 8: Whole corpus projected using multi-CBIR output documents as control points for DT3.

we ranked them extracting the content from their abstracts and performing a string matching algorithm. For this, we employ the terms *"face rendering"*, and *"volumetric human people"* to compute their similarities with the abstracts, respectively.

Table 3: Proposed activities, questions, and group distribution to user evaluation.

	Activity Target	**Question**	**DRIFT**	**List**
A1	Identify a particular group of documents	How many and which documents use human faces for rendering?	GR1	GR2
A2		How many and which documents address volumetric representations of human body?	GR2	GR1

We allow the users to choose any image from DT4 as input to the CBIR component, and to freely define the number (up

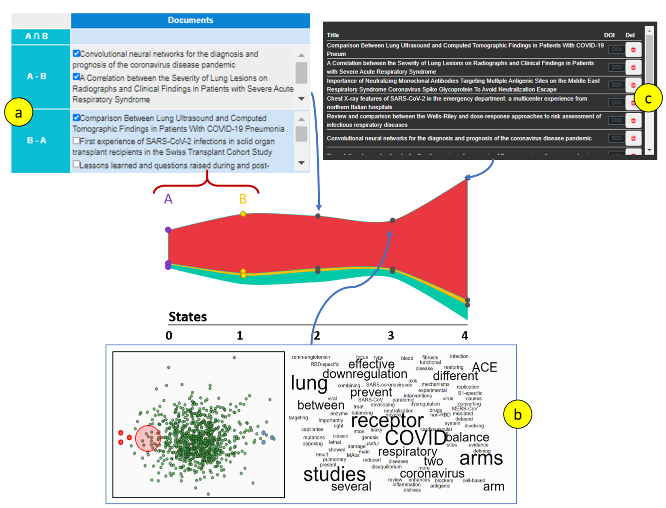
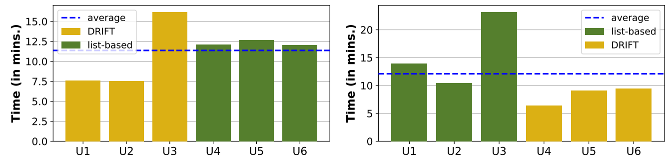


Fig. 9: Displaying selections as a result of exploration in our case study on DT3: (a) we compared the selections of the states 0 and 1 using the modal, and we obtained the state 2, (b) we performed a selection near the red points, and it is saved as the state 3, and (c) we compared states 2 and 3 generating a list of 9 papers related to respiratory infections.

to five) of inputs that they deem necessary to achieve the goal. The users were informed of the maximum number of inputs as part of the Tool Exposure step. Moreover, to minimize human bias, we randomly displayed all the figures on a 2D board and did not limited the time to select the inputs. This elapsed time has not been considered as part of the evaluation.

This study verified the following hypothesis:

- Users of DRIFT will spend less time answering questions



(a) A1 (b) A2

Fig. 10: Comparing spent times (*in minutes*) by the six users (U1-U6) to accomplish A1 and A2 activities.

that require a global analysis of the corpus, with no significant loss in precision.

We computed three well-known information retrieval measures — *i.e.*, *Precision*, *Recall*, and *F1-score* — to evaluate the relevance of document retrieved. To gain a deeper understanding on information retrieval measures, review the book of Schütze *et al.* [41]. Additionally, we stored the elapsed times taken to accomplish A1 and A2 activities. Results are shown in Table 4, for each user and activity. Here, one observes that in A1, users of the list-based interface obtained a significantly lower performance in terms of all measures. Note that the difference between the best precision value for list-based and the worst for DRIFT is close to 0.22, and even that the user U1 performs perfectly the test, obtaining a precision of 1. However, when we inspect the elapsed times in Fig. 10a, we notice that all times list-based users spent almost the same (close to the average value) to perform this task. On the other hand, we note that two of DRIFT users (U1 and U2) obtained the lowest times for this experiment, except for U3 which spent much more time. In A2 activity, GR1' precision average was decreased while GR2 was increased considerably. For instance, the user U4 — who obtained the poorest precision in A1 — improves its performance obtaining 0.75 of precision in A2. Moreover, inspecting the elapsed times in Fig. 10b, one can rapidly notice that DRIFT' users obtained the three lowest times for this activity. The lowest row in Table 4 summarizes both activities by the average calculation.

Finally, to check for statistical significance of the differences found between DRIFT and the list-based approach, we employ a t-test with a 5 percent level ($\alpha = 0.05$). For precision values, we obtained a two-tailed p-value equal to 0.0023, which is considered to be statistically significant. These results confirm our initial hypothesis.

7. User Feedback

After conducting the procedure, all the experienced researchers gave feedback and comments.

User 1 stated: *"The proposal appeals to me since it is designed to search by images combined with components, creating interaction between the elements. I found several papers related to my objective. I see that it could be overwhelmed with some components, and to use it, an explanation is required. It is a helpful tool, and it made my search easier."*

User 2 stated: *"The criteria of searching by image is helpful because when I start a bibliographic review, I usually do not read the entire abstract. I consider that in subjects where images are essential, such as computer graphics, the search for the proposed methodology is handy. In the projection, I'm fond of arranging the points and finding articles without reading the abstract. In text-search, you need prior knowledge to be able to search using keywords, it is not interactive, but it is helpful for research topics in general."*

User 3 stated: *"I had always started my search using traditional methods (by textual search), so your proposal seems very interesting. The first part reminds me of Pinterest when searching for images. There are some things to refine in the projection, it may be to use only the abstract, but with the visual support of the images, I was able to identify suitable papers and store them in my selections. The word cloud helped me to orient myself well in my research. I am satisfied with the papers that I found."*

User 4 stated: *"The work is interesting; using images speeds up the work. I like selecting papers in the projection because I could identify groups with similarities in a graphical way. There is greater precision when combining images and text."*

User 5 stated: *"I found the combination of images and text interesting using analytical and visual search elements. When I look for papers in my area, as visualization topics, I am guided by the images because they are essential for finding the documents of my interest. To do a systematic review, I usually save the papers and classify them as I read them; with DRIFT, I could organize them before reading the abstracts based on the images."*

User 6 stated: *"It is a tool that saves time for the researcher by including the images in the search. I am really into it. I was guided by both the images and their titles to select the articles. However, it depends on the topic to be researched to get articles associated with a search successfully. That is, specific themes do not use images. On the other hand, I enjoy combining my selections and group papers without reading the abstract."*

We received positive feedback from the users, including significant opportunities for our future work, *e.g.*, focus on specific fields of study that deal with a high number of images, and plug it into institutional repositories for exploring theses/dissertations. The researchers showed different behavior in the learning curve for using DRIFT, essentially characterized by his/her background in using this type of visual analytic system. However, all of them described the use of our tool as beneficial for its work. This fact points out the usefulness of our methodology and tool.

8. Discussion and Limitations

The described design and case studies clearly show that our approach provides an efficient alternative for exploring and analyzing extensive collections of scientific documents. Our implementation into the web context aims to introduce a new paradigm into digital libraries' exploration. In that way, we allow analysts to extract insights while mitigating the overwork to establish mental relationships among documents and the excessive time consumption by abstracts reading. DRIFT starts the analytical process with a collection accurately filtered by

Table 4: Measures values (*Precision*, *Recall* and *F1-score*) and in-detail elapsed times (in minutes) obtained by the six participants (U1-U6) of our study after perform the activities (A1, A2).

	list-based				**DRIFT**			
		Precision	**Recall**	**F1-score**		**Precision**	**Recall**	**F1-score**
A1	**U4**	0.200	0.200	0.200	**U1**	1.000	0.600	0.750
	U5	0.286	0.400	0.333	**U2**	0.800	0.800	0.800
	U6	0.333	0.200	0.250	**U3**	0.556	1.000	0.714
A2	**U1**	0.750	0.375	0.500	**U4**	0.750	0.094	0.167
	U2	0.278	0.625	0.385	**U5**	0.750	0.094	0.167
	U3	0.250	0.625	0.357	**U6**	0.700	0.088	0.156
Average		0.349	0.404	0.338		0.759	0.446	0.459

image and textual content, covering a higher number of interesting articles. Then, the complementary resources provide a quick overview of our collection's main topics and metadata, storing this selection to be managed according to the analyst's needs. Finally, this process can be performed several times, allowing us to retrieve and combine previous selections until we reach the desired collection. These features enhance the use of resources in terms of time and effort for compiling a broader and more accurate set of required documents, avoiding the one-by-one review as digital libraries currently have us accustomed to.

The novel combination of multiple-CBIR and multidimensional projection represents a flexible and powerful mechanism to gather image and textual features in a methodology for document exploration. However, the feature extraction step dramatically impacts the whole process of analysis. It is crucial to have a valuable set of features describing images and texts to help us improve the accuracy of searches. Note that not all articles available in digital libraries have images; in these cases, DRIFT only uses textual information since those documents cannot interact with any CBIR component.

We identified some essential considerations in the textual processing step. First, the text size is a decisive factor to consider when using TFIDF or a word embedding model. For instance, the TFIDF model has better accuracy for abstracts than word embedding models, such as doc2vec. However, it employs a large amount of memory to process features since values are highly sparse on vector representation. On the other hand, word embedding models performed better on short texts like titles because the semantic model allows managing the short amount of information while at the same time optimizing memory usage. For a deeper comparison of these two strategies, review Meijer *et al.* [42]. DRIFT makes these two approaches available, allowing the analyst to select the one according to the previously discussed dataset features.

Our implementation visually illustrates the states to be queried, filtered, and combined. It relies on a streamgraph-based plot that performs an advisory role. However, it is not entirely appropriate when document selections are unbalanced, *i.e.*, collections with few elements can be challenging to visualize. On the other side, lower sections of the graph can help reveal outliers. Moreover, it allows us to stack more attributes to visualize simultaneously, and user exploration, *e.g.*, number of reads/downloads or average h-index from the entire collection.

A significant weakness in our prototype concerns the set of operations over states supported. In DRIFT, we only allow the analyst to compare intersections and differences from two states visually. However, if an analyst needs to reach more than two, it will force multiple pair comparisons. For instance, for three states, it will examine the first and second, then the second and third, and lastly, third and first. We intend to implement more operations to improve the analysis experience, *e.g.*, reordering or altering states.

9. Conclusion

In this work, we present DRIFT, a novel visual analytic tool for analyzing scientific literature collection. It comprises multiple linked components such as content-based image retrieval, multidimensional projection, frequency histograms, word clouds, and a streamgraph. We extend the previous version of this paper [4] by adding a new case study, improvements in textual processing, and deeply detailing the user evaluation process. The proposed method is fully interactive, intuitive for analysts aiming to extract subsets of documents according to its requirements. Moreover, it proposes a new paradigm that conciliates both image and textual features into a continuous feedback process. Furthermore, we implemented DRIFT in a web-based environment with the future envision to plug it into a digital library. We demonstrate the usefulness of our methodology in three detailed case studies and user evaluation. Results show that our approach is an attractive method for analyzing multiple types of documents.

Acknowledgment

This work was supported by CONCYTEC-Peru through its executing unit ProCiencia (grant #419-2019), Universidad Católica San Pablo, CNPq-Brazil (grants #303552/2017-4, #312483/2018-0), São Paulo Research Foundation (FAPESP)-Brazil (grant #2013/07375-0) and Getulio Vargas Foundation. The views expressed are those of the authors and do not reflect the official policy or position of the São Paulo Research Foundation.

References

[1] Gomez-Nieto, E, Casaca, W, Motta, D, Hartmann, I, Taubin, G, Nonato, LG. Dealing with multiple requirements in geometric arrangements. IEEE Transactions on Visualization and Computer Graphics 2016;22(3):1223–1235.

[2] Teevan, J, Cutrell, E, Fisher, D, Drucker, SM, Ramos, G, André, P, et al. Visual snippets: summarizing web pages for search and revisitation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009, p. 2023–2032.

[3] Dörk, M, Carpendale, S, Collins, C, Williamson, C. Visgets: Coordinated visualizations for web-based information exploration and discovery. IEEE Transactions on Visualization and Computer Graphics 2008;14(6):1205–1212.

[4] Pocco, X, Poco, J, Viana, M, de Paula, R, Nonato, LG, Gomez-Nieto, E. Drift: A visual analytic tool for scientific literature exploration based on textual and image content. In: 2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). 2021, p. 136–143.

[5] Federico, P, Heimerl, F, Koch, S, Miksch, S. A survey on visual approaches for analyzing scientific literature and patents. IEEE Transactions on Visualization and Computer Graphics 2017;23(9):2179–2198.

[6] Abbasi, A, Altmann, J. On the correlation between research performance and social network analysis measures applied to research collaboration networks. In: System Sciences (HICSS), 2011 44th Hawaii International Conference on. 2011, p. 1–10.

[7] Rodriguez, MA, Pepe, A. On the relationship between the structural and socioacademic communities of a coauthorship network. Journal of Informetrics 2008;2(3):195 – 201.

[8] Morel, CM, Serruya, SJ, Penna, GO, Guimarães, R. Co-authorship network analysis: a powerful tool for strategic planning of research, development and capacity building programs on neglected diseases. PLoS Negl Trop Dis 2009;3(8):e501.

[9] Matejka, J, Grossman, T, Fitzmaurice, G. Citeology: Visualizing paper genealogy. In: CHI '12 Extended Abstracts on Human Factors in Computing Systems. CHI EA '12; New York, NY, USA: Association for Computing Machinery. ISBN 9781450310161; 2012, p. 181–190.

[10] Tarnavsky, A, Smolyansky, E, Itay, K. Connected Papers tool. https://www.connectedpapers.com; 2021.

[11] Liu, S, Chen, C, Ding, K, Wang, B, Xu, K, Lin, Y. Literature retrieval based on citation context. Scientometrics 2014;101(2):1293–1307.

[12] Yan, E, Ding, Y. Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. Journal of the American Society for Information Science and Technology 2012;63(7):1313–1326.

[13] He, J, Ping, Q, Lou, W, Chen, C. Paperpoles: Facilitating adaptive visual exploration of scientific publications by citation links. Journal of the Association for Information Science and Technology 2019;70(8):843–857.

[14] Berger, M, McDonough, K, Seversky, LM. cite2vec: Citation-driven document exploration via word embeddings. IEEE transactions on visualization and computer graphics 2016;23(1):691–700.

[15] Zhang, Y, Ma, Q. Doccit2vec: Citation recommendation via embedding of content and structural contexts. IEEE Access 2020;8:115865–115875.

[16] Dunne, C, Shneiderman, B, Gove, R, Klavans, J, Dorr, B. Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. J Am Soc Inf Sci Technol 2012;63(12):2351–2369.

[17] Beck, F, Koch, S, Weiskopf, D. Visual analysis and dissemination of scientific literature collections with survis. IEEE Transactions on Visualization and Computer Graphics 2016;22(1):180–189.

[18] Pagliosa, P, Martins, RM, Cedrim, D, Paiva, A, Minghim, R, Nonato, LG. Mist: Multiscale information and summaries of texts. In: 2013 XXVI Conference on Graphics, Patterns and Images. 2013, p. 91–98.

[19] Paulovich, FV, Pinho, R, Botha, CP, Heijs, A, Minghim, R. Pex-web: Content-based visualization of web search results. In: 2008 12th International Conference Information Visualisation. IEEE; 2008, p. 208–214.

[20] Choo, J, Lee, C, Clarkson, E, Liu, Z, Lee, H, Chau, DHP, et al. Visirr: Interactive visual information retrieval and recommendation for large-scale document data. Tech. Rep.; Georgia Institute of Technology; 2013.

[21] Wu, S, Zhao, Y, Parvinzamir, F, Ersotelos, NT, Wei, H, Dong, F. Literature explorer: effective retrieval of scientific documents through nonparametric thematic topic detection. The Visual Computer 2019;:1–18.

[22] Deserno, TM, Antani, S, Rodney Long, L. Content-based image retrieval for scientific literature access. Methods of Information in Medicine 2009;48(4):371–380.

[23] Müller, H, Foncubierta-Rodríguez, A, Lin, C, Eggel, I. Determining the relative importance of figures in journal articles to find representative images. In: Medical Imaging 2013: Advanced PACS-based Imaging Informatics and Therapeutic Applications; vol. 8674. International Society for Optics and Photonics; 2013, p. 86740I.

[24] Zhai, A, Kislyuk, D, Jing, Y, Feng, M, Tzeng, E, Donahue, J, et al. Visual discovery at pinterest. In: Proceedings of the 26th International Conference on World Wide Web Companion. WWW '17 Companion; Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2017, p. 515–524.

[25] Yang, F, Kale, A, Bubnov, Y, Stein, L, Wang, Q, Kiapour, H, et al. Visual search at ebay. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017, p. 2101–2110.

[26] Zhang, Y, Pan, P, Zheng, Y, Zhao, K, Zhang, Y, Ren, X, et al. Visual search at alibaba. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18; New York, NY, USA: Association for Computing Machinery; 2018, p. 993–1001.

[27] Felizardo, KR, Salleh, N, Martins, RM, Mendes, E, MacDonell, SG, Maldonado, JC. Using visual text mining to support the study selection activity in systematic literature reviews. In: Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement. ESEM '11; Washington, DC, USA: IEEE Computer Society. ISBN 978-0-7695-4604-9; 2011, p. 77–86.

[28] Felizardo, KR, Andery, GF, Paulovich, FV, Minghim, R, Maldonado, JC. A visual analysis approach to validate the selection review of primary studies in systematic reviews. Information and Software Technology 2012;54(10):1079–1091.

[29] Bergström, P, Atkinson, DC. Augmenting the exploration of digital libraries with web-based visualizations. In: Digital Information Management, 2009. ICDIM 2009. Fourth International Conference on. 2009, p. 1–7.

[30] Chou, JK, Yang, CK. Papervis: Literature review made easy. In: Computer Graphics Forum; vol. 30. Wiley Online Library; 2011, p. 721–730.

[31] Otsu, N. A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics 1979;9(1):62–66.

[32] Gomez-Nieto, E, Roman, FS, Pagliosa, P, Casaca, W, Helou, ES, de Oliveira, MCF, et al. Similarity preserving snippet-based visualization of web search results. IEEE Transactions on Visualization and Computer Graphics 2014;20(3):457–470.

[33] Luhn, HP. The automatic creation of literature abstracts. IBM Journal of research and development 1958;2(2):159–165.

[34] Le, Q, Mikolov, T. Distributed representations of sentences and documents. In: International conference on machine learning. PMLR; 2014, p. 1188–1196.

[35] Rehurek, R, Sojka, P. Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 2011;3(2).

[36] Krizhevsky, A, Sutskever, I, Hinton, GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 2012, p. 1097–1105.

[37] Joia, P, Coimbra, D, Cuminato, JA, Paulovich, FV, Nonato, LG. Local affine multidimensional projection. IEEE Transactions on Visualization and Computer Graphics 2011;17(12):2563–2571.

[38] Nonato, LG, Aupetit, M. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. IEEE Transactions on Visualization and Computer Graphics 2018;25(8):2650–2673.

[39] Maaten, Lvd, Hinton, G. Visualizing data using t-sne. Journal of machine learning research 2008;9(Nov):2579–2605.

[40] McInnes, L, Healy, J, Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426 2018;.

[41] Schütze, H, Manning, CD, Raghavan, P. Introduction to information retrieval; vol. 39. Cambridge University Press Cambridge; 2008.

[42] Meijer, H, Truong, J, Karimi, R. Document embedding for scientific articles: Efficacy of word embeddings vs tfidf. arXiv preprint arXiv:210705151 2021;.