

Online Resource 2

Transcriptus Access and JoinFinder Interface Tutorial

Omer Ventura

Daria Vasyutinsky-Shapira

Nachum Dershowitz

Supplementary material for

JoinFinder: A Multi-Modal Architecture for Historical Manuscript Page Identification

International Journal on Document Analysis and Recognition (IJDAR)

School of Computer Science and AI, Tel Aviv University

Corresponding author: Omer Ventura, omerventura@mail.tau.ac.il

1 Introduction

This document provides instructions for accessing and using the interactive JoinFinder interface available through Transcriptus. The interface exposes retrieval results generated from the manuscript representations described in the accompanying paper and provides tools for inspecting manuscript images, candidate joins, and graph-based manuscript relations.

The interface is intended as an exploratory tool. Retrieved neighbors and graph connections should be interpreted as computationally generated candidates for scholarly inspection rather than as automatically confirmed manuscript joins.

2 Accessing Transcriptus

Open the following address in a web browser:

<https://transcriptus.org/>

If a username and password are requested, use:

Username: reader

Password: bb13805ab2a02ae83cfdca319b1dc11a

After entering Transcriptus, the JoinFinder functionality can be accessed through the navigation panel. Depending on the size of the browser window, the navigation panel may appear as a collapsible sidebar.

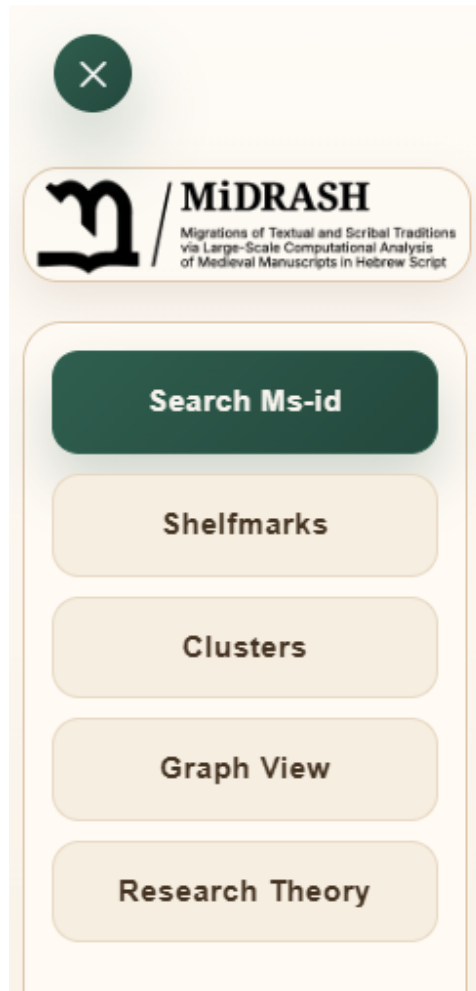


Figure 1: Transcriptus navigation panel. The interface provides entry points for manuscript-ID search, shelfmark-based exploration, clusters, the graph view, and accompanying research information.

3 Searching for a Manuscript

Select **Search Ms-id** from the navigation panel. The manuscript search page contains a field in which a manuscript identifier (`manuscript_id`) can be entered.

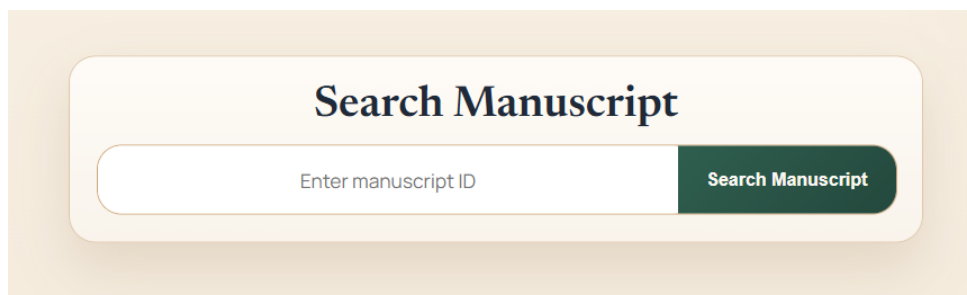


Figure 2: Manuscript-ID search interface. Enter a `manuscript_id` and press *Search Manuscript*.

Enter the desired manuscript identifier and press **Search Manuscript**. The resulting manuscript view displays the images associated with the selected catalog manuscript. Images can be opened

for closer inspection, and the gallery controls can be used to move through larger manuscripts page by page.

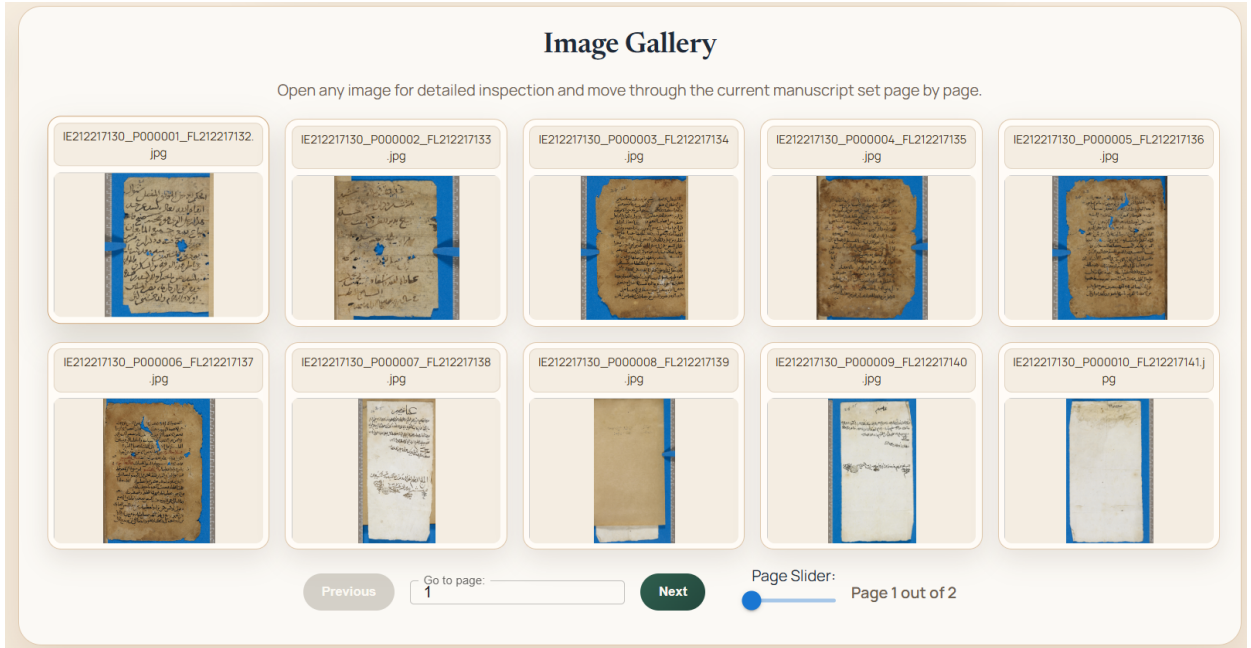


Figure 3: Example image gallery for a selected manuscript. The gallery displays the available manuscript images and provides controls for moving through additional pages.

4 Inspecting Suggested Joins

For a selected manuscript, JoinFinder can display highly ranked neighboring fragments from the learned embedding space. Each candidate is shown together with its retrieval rank and similarity value.

The interface presents the query fragment and its retrieved neighbor side-by-side. The associated manuscript identifiers, image identifiers, libraries, and shelfmarks are shown below the images where available.

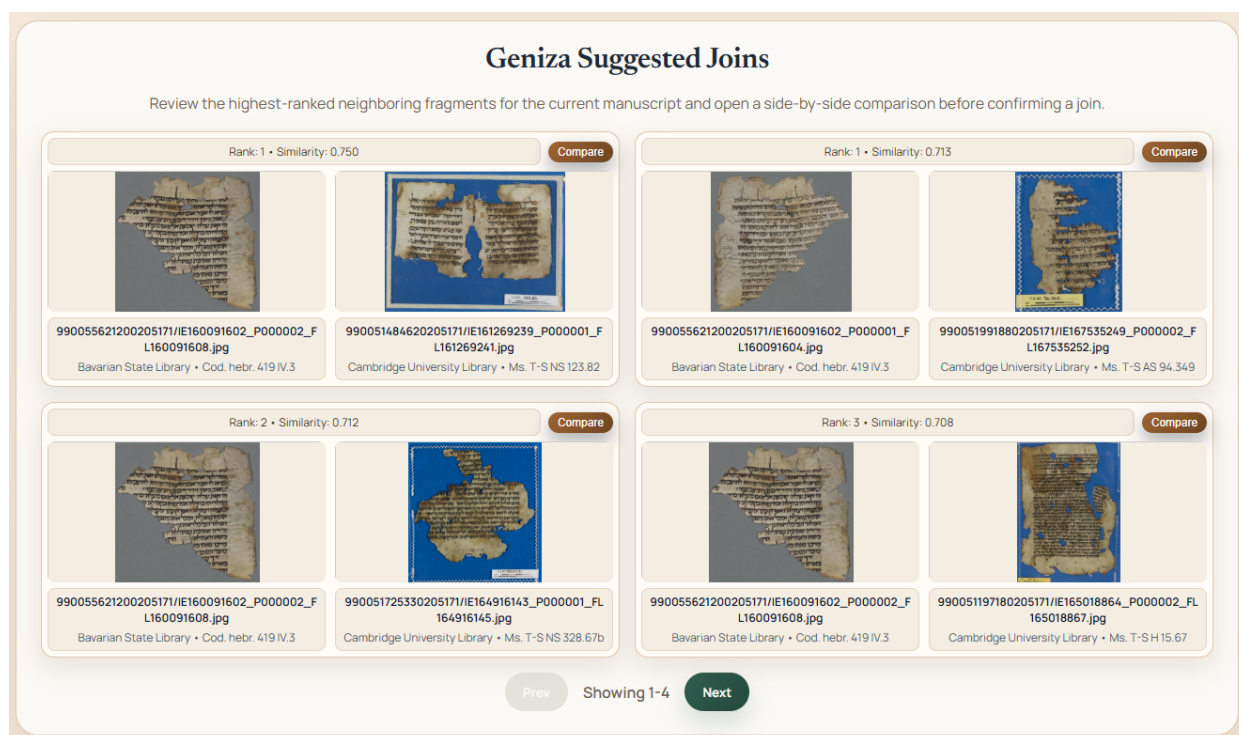


Figure 4: Example JoinFinder suggested-joins view. Candidate neighboring fragments are ranked by similarity and can be opened for side-by-side inspection using the *Compare* control.

The **Compare** button can be used to inspect a candidate pair more closely. The **Prev** and **Next** controls allow navigation through additional retrieved candidates.

In some cases, multiple pages belonging to the same catalog manuscript may appear among the retrieved relations. Such results are useful for examining the behavior of the learned representation but should be distinguished from cross-manuscript candidate joins.

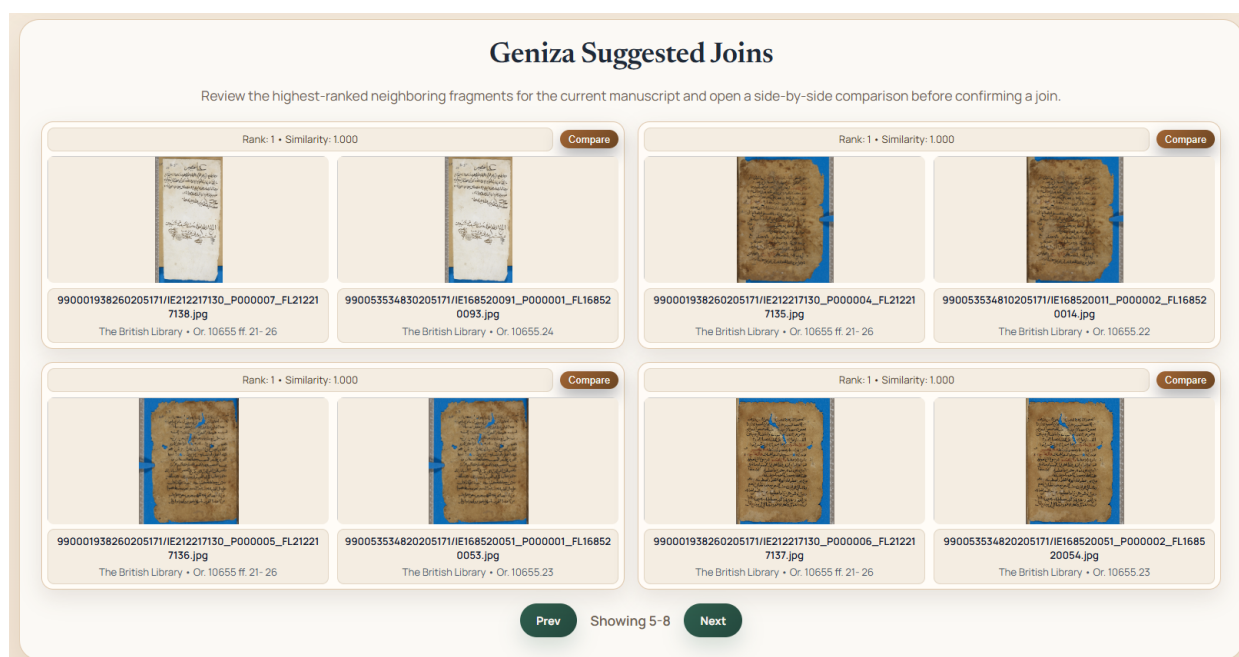


Figure 5: Example retrieval results involving different pages associated with the same catalog manuscript.

5 Unavailable Images

Some manuscript images may not be displayed in the interface. The main reason is collection-specific access and reproduction restrictions: certain holding libraries or institutions do not permit their digitized manuscript images to be displayed through Transcriptus, even when the corresponding manuscript metadata and retrieval relations are available.

In these cases, the interface shows an **Image unavailable** placeholder instead of the manuscript image.



Figure 6: Example of an unavailable image. In most cases, images are not displayed because of collection-specific access or reproduction restrictions. More rarely, an image may be unavailable because the corresponding IIIF request fails.

An unavailable image does not imply that the corresponding manuscript record, embedding, or retrieval relation is missing or invalid. The associated metadata and retrieval result may still be present even when the image itself cannot be shown. In a smaller number of cases, an image may be genuinely unavailable because the external IIIF request is malformed, incomplete, or otherwise fails to resolve.

6 Exploring Manuscript Relations in the Graph View

The **Graph View** provides an interactive visualization of manuscript-level relations derived from the retrieved page-level connections.

6.1 Loading the graph

The graph contains a large number of manuscript nodes and relations and may require some time to initialize. After opening the graph view, wait for the visualization to finish loading before interacting with it.

Once loaded, the graph can be explored interactively. The user can zoom in and out to move between a broad view of manuscript communities and a detailed view of individual nodes and connections.

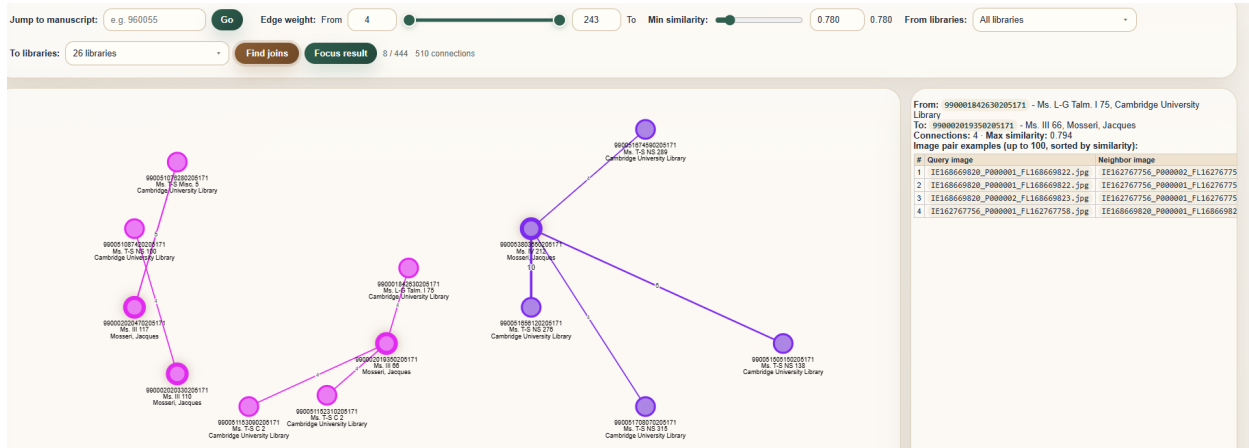


Figure 7: Interactive manuscript graph. Manuscript nodes and the relations between them can be explored after the graph has finished loading. Controls at the top of the interface allow the displayed graph to be filtered.

6.2 Nodes and manuscript identifiers

Each graph node represents a catalog manuscript. At sufficiently high zoom, the node label displays the associated `manuscript_id` together with available catalog information.

Users can click different nodes to inspect the corresponding manuscript and can use the displayed manuscript identifier to relate the graph node back to the manuscript-search interface.

The **Jump to manuscript** field can also be used to locate a specific manuscript identifier directly in the graph.

6.3 Inspecting connections

Edges represent retained relations between manuscript nodes. Numbers displayed along the edges indicate the amount of supporting connection evidence shown by the graph.

Clicking a connection opens information about the two manuscripts associated with that relation. The information panel displays the `manuscript_id` of the manuscripts at both ends of the connection, their available catalog descriptions, the number of supporting connections, the maximum similarity, and examples of the corresponding retrieved image pairs.

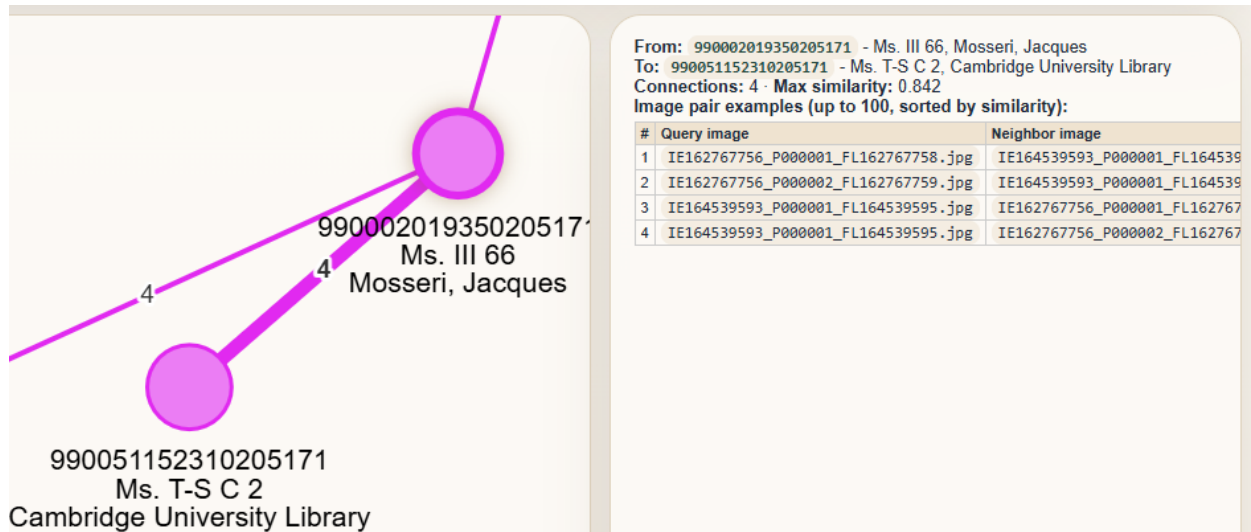


Figure 8: Detailed graph view. Selecting a manuscript relation exposes the manuscript identifiers at both ends of the connection together with supporting image-pair information.

The graph should therefore be used as an exploratory layer: strong or repeated relations can be inspected in greater detail, but a displayed edge should not by itself be interpreted as proof of a physical or codicological join.

6.4 Filtering the graph

The controls above the graph can be used to reduce the displayed set of relations. The interface provides controls for connection weight, minimum similarity, source and destination libraries, manuscript lookup, and focusing on selected results.

These filters are particularly useful when searching for cross-library candidate relations, since the graph can otherwise contain a large number of nodes and edges.

7 Recommended Exploration Workflow

A typical JoinFinder exploration can be performed as follows:

1. Open <https://transcriptus.org/> and authenticate if requested.
2. Select **Search Ms-id** and enter a known `manuscript_id`.
3. Inspect the manuscript images in the image gallery.
4. Examine the ranked suggested joins and use **Compare** to inspect promising candidate pairs.
5. Open **Graph View** and wait for the graph to finish loading.
6. Zoom into the relevant graph region and inspect manuscript nodes.
7. Select edges to examine the manuscript identifiers and supporting image-pair relations associated with them.
8. Use similarity, edge-weight, and library filters to narrow the graph when searching for specific types of cross-manuscript or cross-library relations.

8 Interpretation

JoinFinder is designed to prioritize potentially relevant manuscript relations for expert review. Similarity values, ranked neighbors, and graph connections represent evidence produced by the learned retrieval representation; they are not automatic scholarly determinations of manuscript identity or physical joining.

Candidate relations should therefore be evaluated together with manuscript imagery, catalog metadata, textual and paleographic evidence, and relevant domain expertise.