User Manual of the BHL Visual Search Interface

We have developed a web-based visual search interface[1] that enables users to search the Biodiversity Heritage Library (BHL)[2] library using queries consisting of species names (i.e., mammals, flowering plants and birds). The underlying search engine incorporates a term inventory that is used to suggest semantically related terms based on a measure of relatedness to the initial query. With the integrated term inventory, the interface can show users not only documents matching the original query but also those that mention semantically relevant species which have, for example, shared characteristics such as habitat or taxonomic classification. This allows users to expand an initially query with additional terms that are semantically related to the input query, thus increasing the number of relevant documents retrieved by the search engine.

This manual provides a step-by-step guide to the various functionalities of the search interface.

Step 1: Keyword-based search

The first time a user accesses the interface, he/she is presented with the following four graphical control elements (as shown in Figure 1): a) a search box which accepts a user-specified query, b) a search button which is used to submit the query to the search engine c) an 'About' button which displays information about the project and the relevant publication and d) a 'Help' button which summarises the functionalities of the visual search interface and provides a download link to the user manual of the interface (i.e., present document).
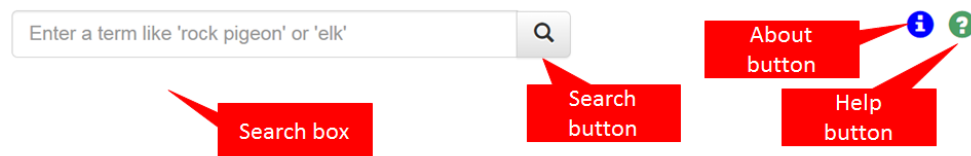

Figure 1. Users can search for a species name using the search box located in the top right panel of the interface

The interface implements a standard keyword-based search engine that retrieves BHL documents containing species names specified by the user. To start using the keyword-based search, enter a species name in the search box which can be found in the top corner of the screen. The search box uses an auto-complete functionality that tries to complete the species name as the user is typing into the box. In the example in Figure 2, the user is typing the word 'rock' and the system recommends different species names that contain the typed word (e.g., 'rock dove').

---

[1] http://nactem10.mib.man.ac.uk/va/MiBio/Search/queryExpansion.html?prot=thumb
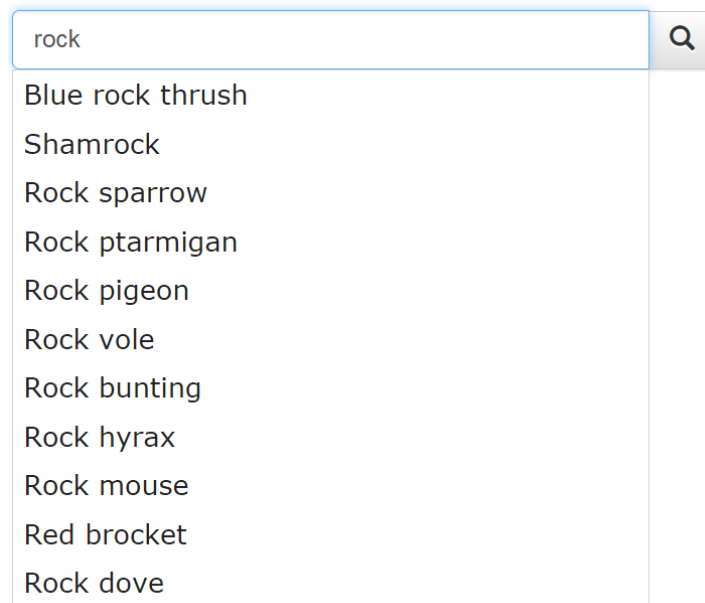[2] http://www.biodiversitylibrary.org/

Figure 2. Auto-completion functionality

In addition to the auto-complete functionality, the search box uses an auto-correct mechanism that can help to correct a misspelled species name (as illustrated in Figure 3).

Figure 3. Auto-correct functionality

To submit your query, simply click on the search icon or select one of the suggested species names. BHL documents matching the user-specified query (i.e., 'rock dove') are listed in the 'Search results' panel which can be found on the right of the screen (just below the search box). For each document, the system displays the following information (illustrated in Figure 4): a) title, b) publisher, c) publication day and d) snippets of the document that contain the query ('rock dove' appears as emboldened text in each snippet). Clicking on a document re-directs the user to the corresponding page in BHL.
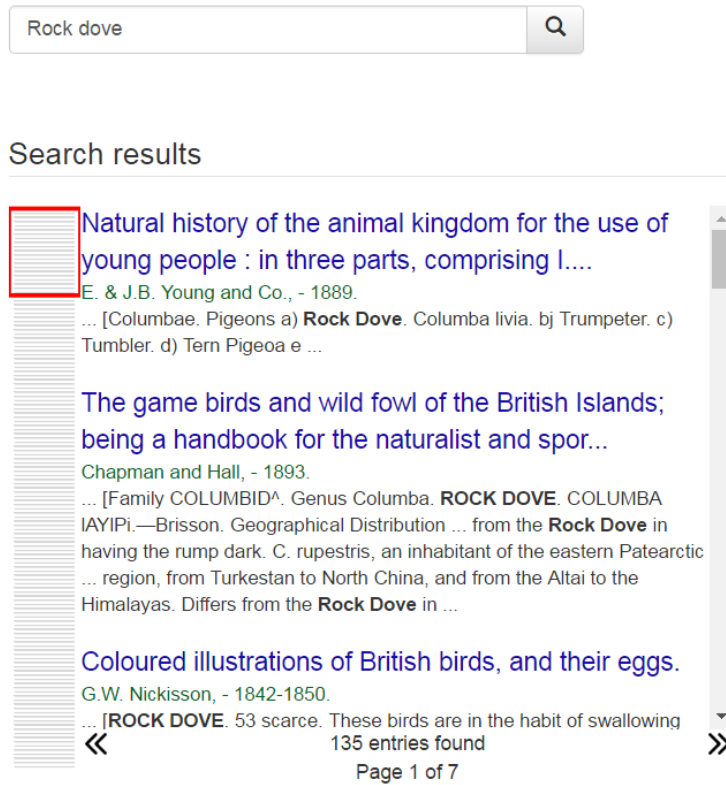
Figure 4. BHL documents matching the input query-keyword

Search results are organised into pages each of which consists of 20 documents. In Figure 5, clicking on the previous/next arrows which can be found on the bottom of the screen allows users to navigate through the different search result pages. The context viewer on the left-hand side shows a zoomed out view of the retrieved list and can be used as an alternative way of navigating through the search result pages.
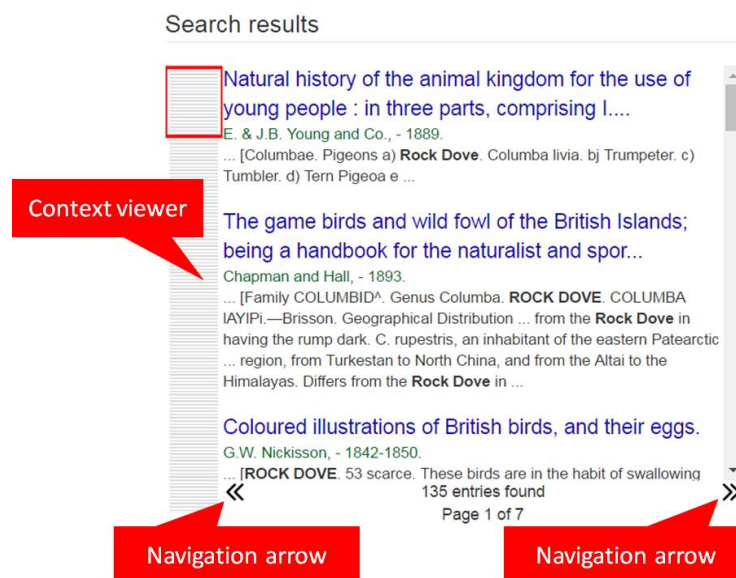

Figure 5. The Navigation arrows and the context viewer allow users to navigate through the different search result pages.

Step 2: Query expansion

The system enables query expansion by automatically suggesting species names which are semantically related to the input query. The recommended species names are displayed on right-hand side of the screen as thumbnails (Figure 6).
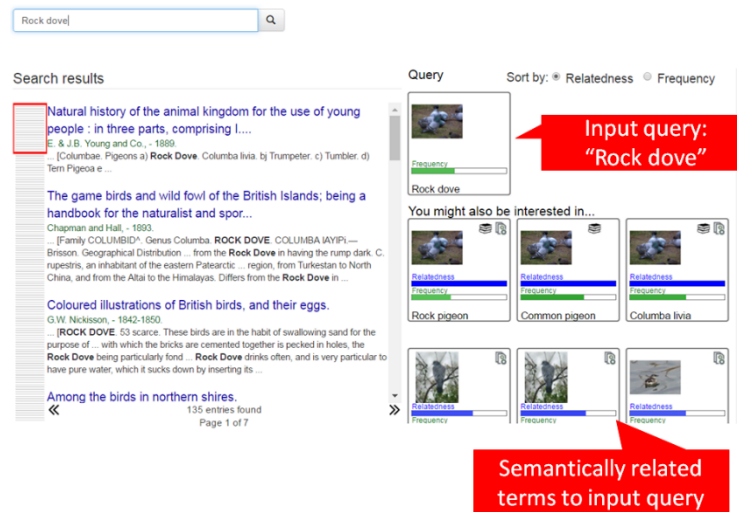

Figure 6. Semantically related terms to an input query are represented as thumbnails.

By default, recommended species names are ranked according to a measure of semantic relatedness to the initial query (Figure 7). Alternatively, users can rank the suggested species names by their frequency, i.e., number of times a species name appear in the BHL corpus.
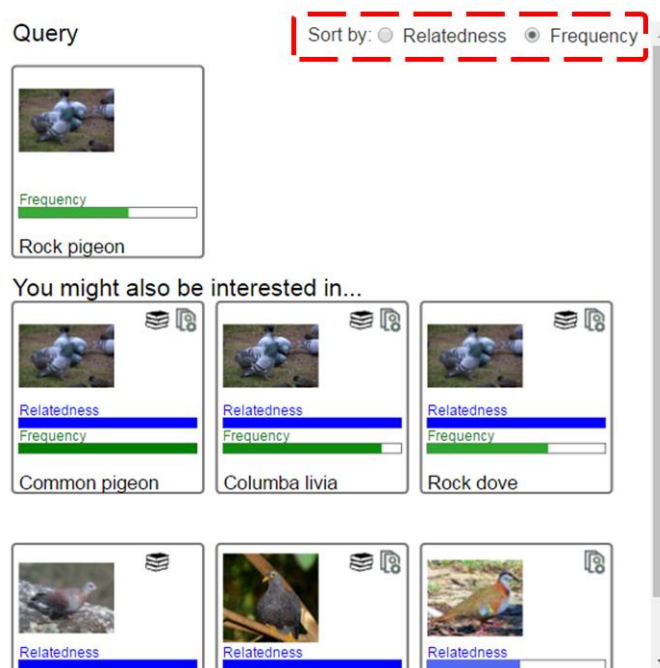

Figure 7. Automatically suggested species names can be ranked by a measure of semantic relatedness to the input query or by their frequency

Each thumbnail (Figure 8) is displayed with two parallel indicator bars, one representing the term's frequency in the BHL corpus and the other its relatedness to the query term. Within the thumbnails, there are also small icons that indicate whether the name exists as a semantically related term in our term inventory, in an external taxonomy or in both.
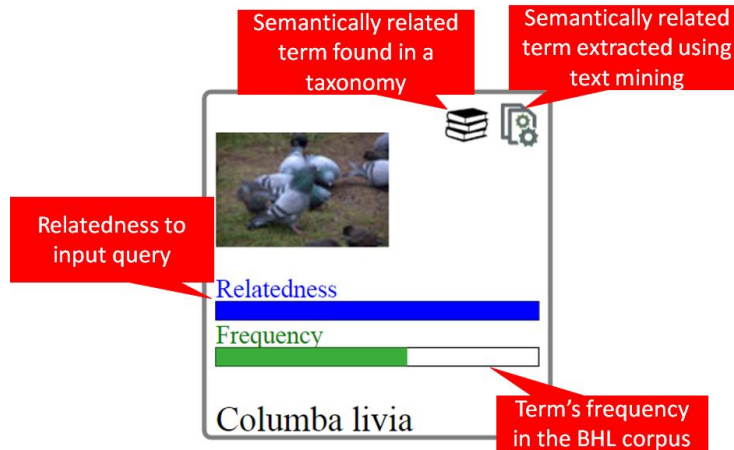
Figure 8. A thumbnail corresponding to a semantically related term to the input query

Clicking on a thumbnail will extend the initial query with an additional term (i.e., species name corresponding to the selected thumbnail). Different query terms are combined using two search operators, namely 'OR' and 'AND' (Figure 9). The 'OR' operator retrieves documents containing any of the query term while the 'AND' operator retrieves documents that contain all query terms. By default the system will use the 'OR' operator which allows the search engine to retrieve a larger number of relevant documents.
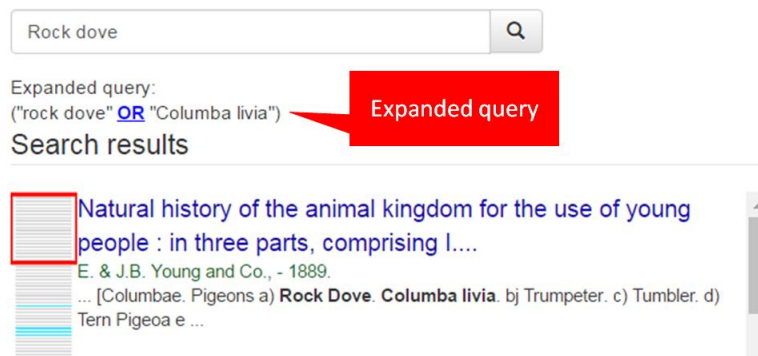

Figure 9. Query expanded with a semantically related term

The selected search operator appears in the expanded query as a hyperlink. Clicking on the hyperlink will toggle between the two operators.

In order to help the user assess how the list of retrieved documents has changed as a result of expanding the query, the system displays the number of entries that are either added or removed from the document list, when the 'OR' or 'AND' operator is used, respectively.
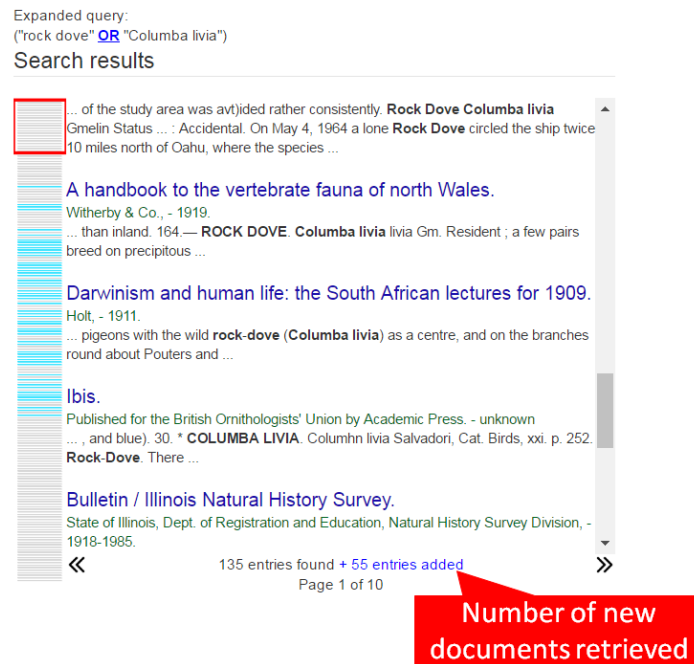
Figure 10. Number of new documents retrieved by the system as a result of expanding the initial, user-specified query (i.e., "rock dove") with an additional species name (i.e., "Columba livia"). The two query species names are combined using the 'OR' operator.

In the case of the 'OR' operator, the system uses different background colours, both in the list of retrieved documents and in the context viewer to determine which documents were retrieved as a result of expanding the query, and which ones were part of the original search result (e.g., documents retrieved according to the expanded query are shown with a light blue background in the search result list and with light blue lines in the context viewer).
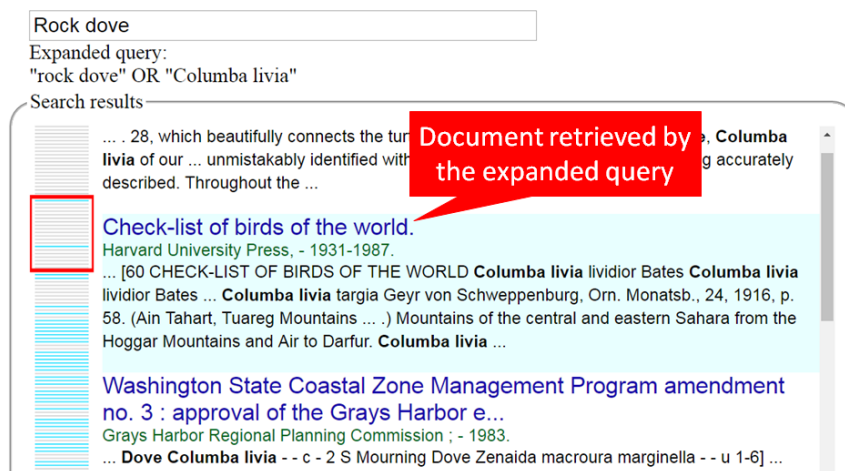

Figure 11. A document retrieved by the expanded query are showing

Documents that are discarded whilst using the 'AND' operator are represented in the context viewer using blank lines (Figure 12). Apparently, with "OR" operator, the number of returned documents is always higher than that by "AND" operator. Another point is that if users expand the query by clicking on high-frequency related terms, the system will return much more documents than those by low-frequency ones.
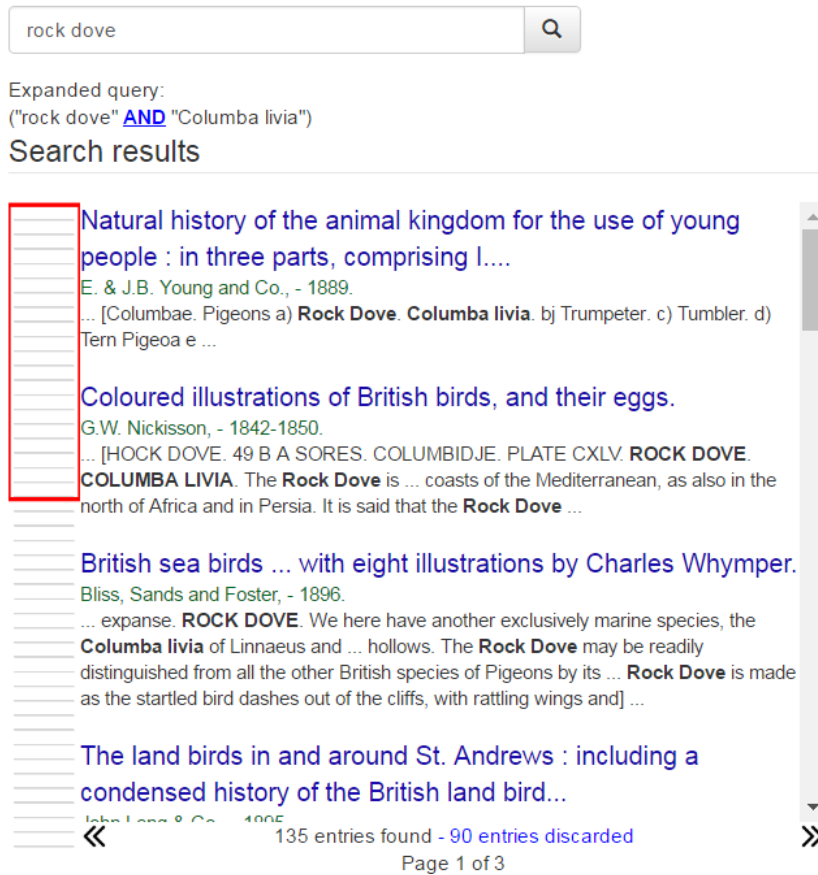
Figure 12. Blank lines in the context viewer represent documents that are discarded as a result of using the 'AND' operator

The interface implements an 'Undo' functionality which can be used to remove a species name from the expanded query (Figure 13). For this, simply click on the thumbnail of a previously selected species name.
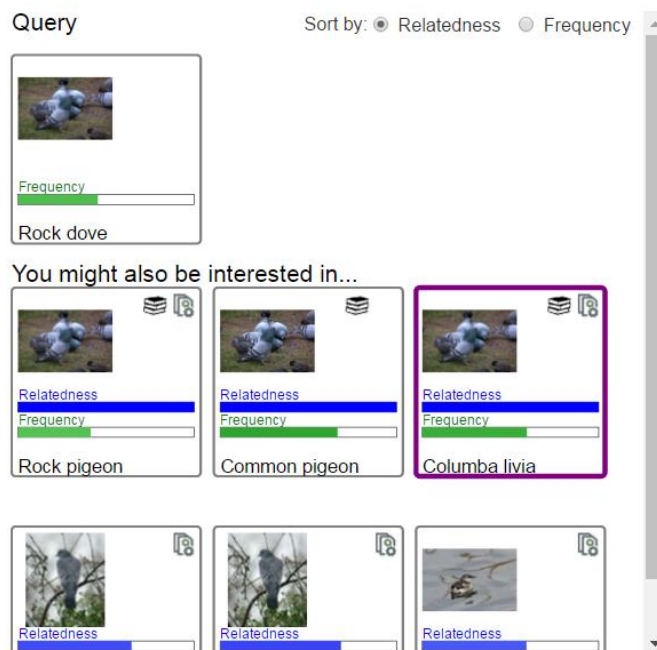

Figure 13. Undo functionality: Clicking on the thumbnail of a previously selected species name will remove the corresponding term from the expanded query

Browser compatibility

The web-based visual search interface is compatible with the following web browsers: a) Google chrome, b) Safari and c) Opera.