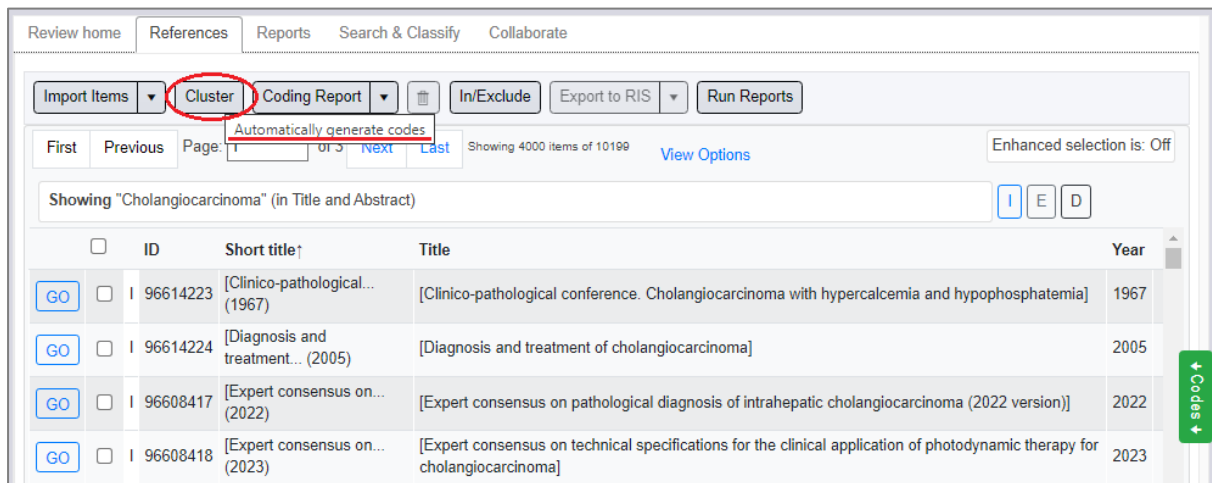


Clustering

Clustering or Auto-coding in EPPI-Reviewer can categorise your references into clusters, based on the text in their abstracts. This automatic document clustering, using text mining, is one way of describing the range of studies you have identified at the click of a button.

Text mining can assist with searching by identifying significant terms in the documents you have already included. Clustering takes this a step further by taking the extracted terms and coding the items into a tree structure, consisting of the extracted terms that can be searched and manipulated. This is done using the Lingo3G clustering engine.

To auto-code references click on the **Cluster** button on the **References** tab.



Review home | References | Reports | Search & Classify | Collaborate

Import Items | **Cluster** | Coding Report | In/Exclude | Export to RIS | Run Reports

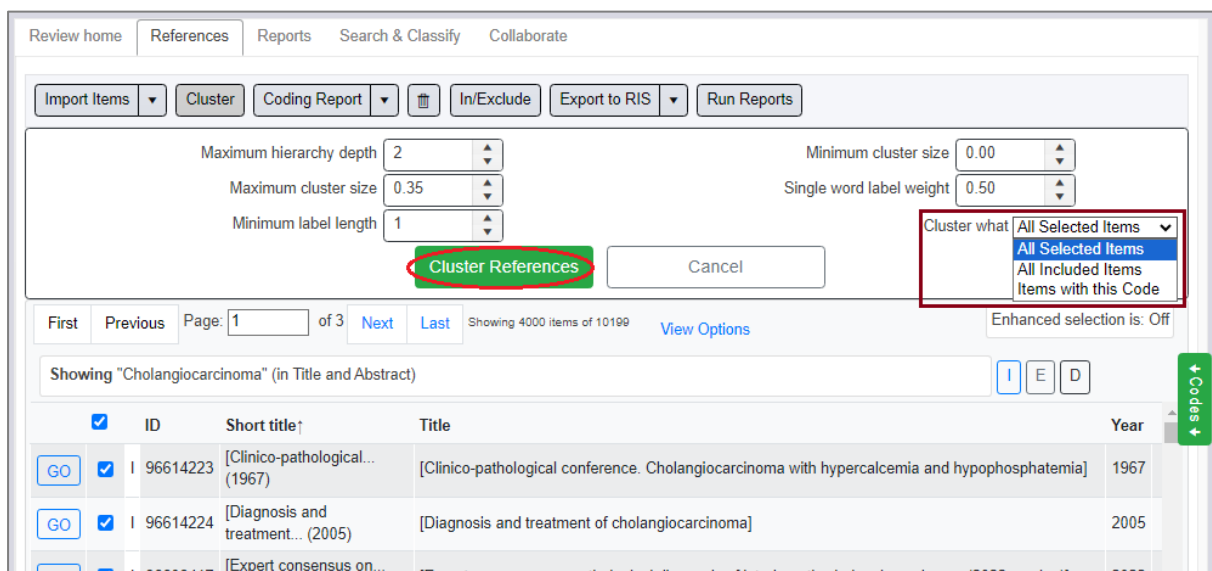
Automatically generate codes

First Previous Page: 1 of 3 Next Last Showing 4000 items of 10199 View Options Enhanced selection is: Off

Showing "Cholangiocarcinoma" (in Title and Abstract) [I] [E] [D]

<input type="checkbox"/>	ID	Short title†	Title	Year
<input type="checkbox"/>	96614223	[Clinico-pathological... (1967)]	[Clinico-pathological conference. Cholangiocarcinoma with hypercalcemia and hypophosphatemia]	1967
<input type="checkbox"/>	96614224	[Diagnosis and treatment... (2005)]	[Diagnosis and treatment of cholangiocarcinoma]	2005
<input type="checkbox"/>	96608417	[Expert consensus on... (2022)]	[Expert consensus on pathological diagnosis of intrahepatic cholangiocarcinoma (2022 version)]	2022
<input type="checkbox"/>	96608418	[Expert consensus on... (2023)]	[Expert consensus on technical specifications for the clinical application of photodynamic therapy for cholangiocarcinoma]	2023

A window will appear where you can set the parameters of the clustering and what items are clustered.



Review home | References | Reports | Search & Classify | Collaborate

Import Items | **Cluster** | Coding Report | In/Exclude | Export to RIS | Run Reports

Maximum hierarchy depth: 2 | Minimum cluster size: 0.00

Maximum cluster size: 0.35 | Single word label weight: 0.50

Minimum label length: 1

Cluster References | Cancel

Cluster what: All Selected Items (dropdown menu)

First Previous Page: 1 of 3 Next Last Showing 4000 items of 10199 View Options Enhanced selection is: Off

Showing "Cholangiocarcinoma" (in Title and Abstract) [I] [E] [D]

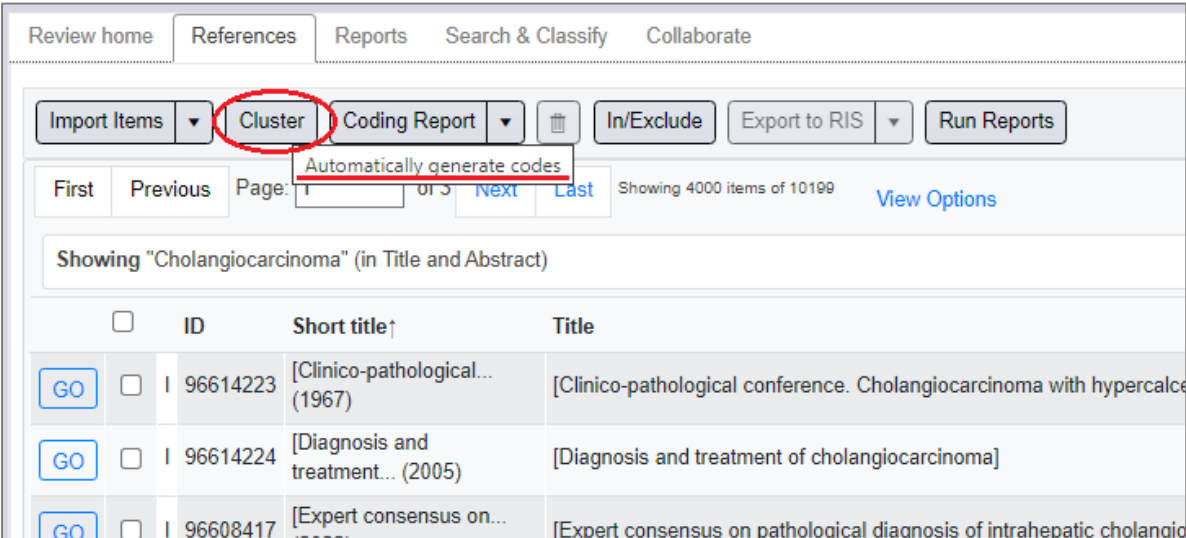
<input checked="" type="checkbox"/>	ID	Short title†	Title	Year
<input checked="" type="checkbox"/>	96614223	[Clinico-pathological... (1967)]	[Clinico-pathological conference. Cholangiocarcinoma with hypercalcemia and hypophosphatemia]	1967
<input checked="" type="checkbox"/>	96614224	[Diagnosis and treatment... (2005)]	[Diagnosis and treatment of cholangiocarcinoma]	2005
<input checked="" type="checkbox"/>	96608417	[Expert consensus on... (2022)]	[Expert consensus on pathological diagnosis of intrahepatic cholangiocarcinoma (2022 version)]	2022

The parameters are -:

- Maximum hierarchy depth
 - o Default 2.0
 - o Determines how deep in the code tree to go
- Minimum cluster size
 - o Default 0.0 (values 0.0 to 1.0)
- Maximum cluster size
 - o Default 0.35 (values 0.0 1.0)
- Single word label weight
 - o Default 0.5
 - o How the minimum label length is weighted.
- Minimum label length
 - o Default 1.0
 - o The minimum number of words in a cluster

You can either select the items to cluster ahead of time, run it on all included items, or run it on items with a particular code.

Once the options have been set as you wish, click on **Cluster References** button.



The screenshot shows a web interface with a navigation bar at the top containing tabs for 'Review home', 'References', 'Reports', 'Search & Classify', and 'Collaborate'. Below the navigation bar is a toolbar with buttons for 'Import Items', 'Cluster', 'Coding Report', 'In/Exclude', 'Export to RIS', and 'Run Reports'. The 'Cluster' button is circled in red. A tooltip 'Automatically generate codes' is visible over the 'Cluster' button. Below the toolbar is a pagination section with 'First', 'Previous', 'Page: 1 of 3', 'Next', and 'Last' buttons, along with the text 'Showing 4000 items of 10199' and a 'View Options' link. The main content area displays a search result for 'Cholangiocarcinoma' (in Title and Abstract) with a table of results. The table has columns for 'ID', 'Short title', and 'Title'. The first three rows are visible:

<input type="checkbox"/>	ID	Short title†	Title
<input type="checkbox"/>	96614223	[Clinico-pathological... (1967)]	[Clinico-pathological conference. Cholangiocarcinoma with hypercalcaemia]
<input type="checkbox"/>	96614224	[Diagnosis and treatment... (2005)]	[Diagnosis and treatment of cholangiocarcinoma]
<input type="checkbox"/>	96608417	[Expert consensus on... (2002)]	[Expert consensus on pathological diagnosis of intrahepatic cholangiocarcinoma]

The newly created codes will appear in the Code Tree on the right of the screen, under a new coding tool called 'Lingo3G clusters'.

The screenshot shows a software interface with a table on the left and a code tree on the right. The table has a 'Year' column and contains the following entries:

	Year
iolocarcinoma with hypercalcemia and	1967
cinoma]	2005
osis of intrahepatic cholangiocarcinoma	2022
ions for the clinical application of oma]	2023
	2019
ocellular carcinoma in cows]	1977
ytic cholangiocarcinoma prognosis ity of cancer cells	2021
r	2006
y cholangitis	2009
he role of vascular resection in the	2013
ocarcinoma	1999
resection of hilar cholangiocarcinoma	2006

The code tree on the right is titled 'Lingo3G clusters' and is highlighted with a red box. It contains the following items:

- ▶ Screen on Title & Abstract
- ▶ References
- ▼ Lingo3G clusters
 - ▶ Intrahepatic Cholangiocarcinoma
 - ▶ Hepatocellular Carcinoma
 - ▶ Expression
 - ▼ Overall Survival
 - Gemcitabine and Cisplatin
 - Tumor
 - Progression-free Survival
 - Biliary Tract Cancer
 - iCCA Patients
 - Locally Advanced
 - Liver Metastasis
 - Neoadjuvant
 - (Other topics)
 - Liver Transplantation
 - Primary Sclerosing Cholangitis
 - Lymph Node
 - Duct Carcinoma
 - Review of the Literature
 - IFN-gamma
 - Biliary Strictures

(If you run the clustering tool again, a new set of codes will be generated under the same default coding tool name, so we suggest renaming these if you intend to try multiple uses of the tool.)

Now that the items are coded with the extracted terms the items can be searched and manipulated similar to any other coded item.

This clustering functionality becomes useful when dealing with search results that are in the range of many thousands. It allows you to quickly categorise, or 'Auto-code' the terms contained in the documents and identify the most relevant documents to your subject.

What you may also find useful is running **Frequency Reports** on the clustering codesets to see how common the terms found are across your references e.g.

The screenshot shows the Lingo 3G interface with the 'Reports' tab selected. The 'Run Reports' section is active, showing a frequency report for 'Intrahepatic Cholangiocarcinoma'. The report table is as follows:

Code	Count
(Other topics)	9
Carcinoma	6
Gemcitabine and Cisplatin	6
Gene Expression	7
iCCA Patients	8
Laparoscopic Liver	2

The right-hand sidebar shows a list of Lingo3G clusters, with 'Intrahepatic Cholangiocarcinoma' highlighted. A red circle highlights the 'With this Code' button in the top right corner.

Crosstab reports could also be run to see where items have codes from multiple codesets within the Lingo 3G coding tool.

The screenshot shows the Lingo 3G interface with the 'Reports' tab selected. The 'Run Reports' section is active, showing a crosstab report for 'Intrahepatic Cholangiocarcinoma'. The report table is as follows:

	Gemcitabine and Cisplatin	Tumor	Progression-free Survival	Biliary Tract Cancer	iCCA Patients	Locally Advanced	Liver Metastasis	Neoadjuvant
iCCA Patients	2	1	2	0	3	1	0	0
Surgical Resection	1	2	1	0	1	1	1	1
Patients Receiving	3	0	1	0	2	2	1	2
Gene Expression	1	0	1	0	1	0	0	0
Progression-free Survival	2	1	3	1	2	1	0	0