


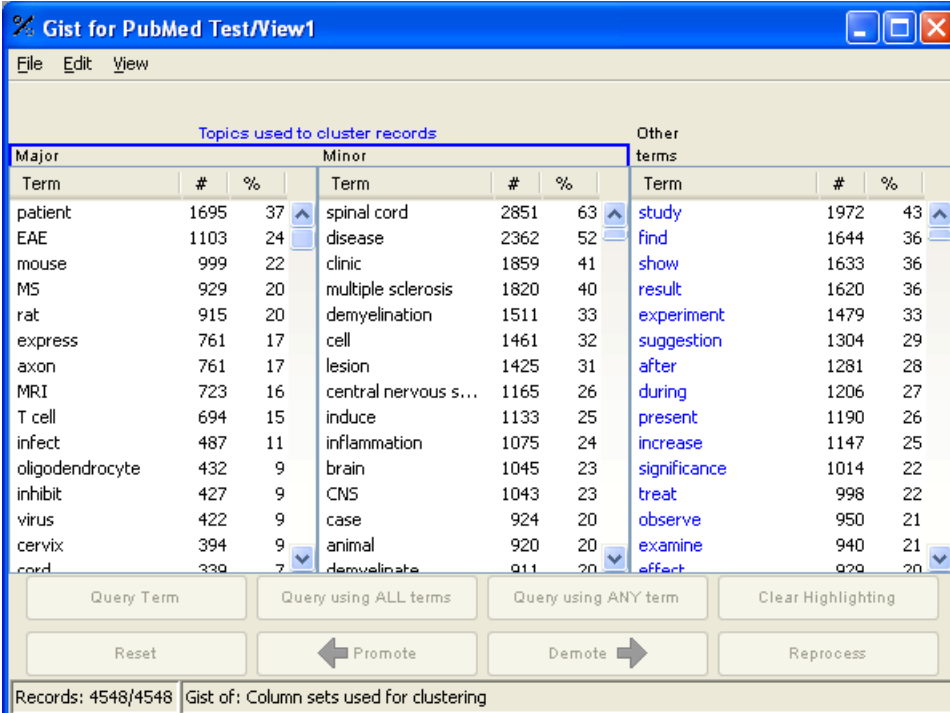


Getting started with OmniViz

Understanding the Gist of a Document Set

The Gist Tool is used to get a quick overview of the vocabulary contained in selected documents, and to control how those topics influence clustering of documents. You can also query for documents containing topics of interest.

1. Open OmniViz. In the **Projects** window, navigate to your data set and **View** of interest by clicking on the plus signs. Double click on a View to open up the **Galaxy** visualization.
2. Select all documents in the Galaxy by pressing **Ctrl+A**.
3. Open up the **Gist** tool by selecting it from the **Tools** menu in the **OmniViz Desktop** (or by clicking on the Gist icon ).



Topics used to cluster records						Other terms		
Major			Minor			Other terms		
Term	#	%	Term	#	%	Term	#	%
patient	1695	37	spinal cord	2851	63	study	1972	43
EAE	1103	24	disease	2362	52	find	1644	36
mouse	999	22	clinic	1859	41	show	1633	36
MS	929	20	multiple sclerosis	1820	40	result	1620	36
rat	915	20	demyelination	1511	33	experiment	1479	33
express	761	17	cell	1461	32	suggestion	1304	29
axon	761	17	lesion	1425	31	after	1281	28
MRI	723	16	central nervous s...	1165	26	during	1206	27
T cell	694	15	induce	1133	25	present	1190	26
infect	487	11	inflammation	1075	24	increase	1147	25
oligodendrocyte	432	9	brain	1045	23	significance	1014	22
inhibit	427	9	CNS	1043	23	treat	998	22
virus	422	9	case	924	20	observe	950	21
cervix	394	9	animal	920	20	examine	940	21
cord	339	7	demyelinate	911	20	effect	929	20

Query Term Query using ALL terms Query using ANY term Clear Highlighting

Reset Promote Demote Reprocess

Records: 4548/4548 Gist of: Column sets used for clustering

The Gist window shows the **Major**, **Minor** and **Other** terms of the selected documents. The Major terms (up to 200) are the words or phrases that are most relevant to discriminating one document from another; these are the terms used in clustering the documents. The

Minor terms (up to 2000) are a lesser set of discriminating terms—they are used indirectly in clustering by providing context for the Major terms. The Other terms consist of every other word/phrase in the document set. # indicates the number of documents in which a term occurs. Look through the Major and Minor terms to get an idea of the subject matter of the document set.

4. To gain insight into how document clusters were formed, select each cluster in turn by clicking on cluster centroids in Galaxy. For each cluster, look at the top Major terms and note the percent of selected documents they occur in. The more high percentage terms you see, the stronger the clustering. You will probably see terms at the top of the list for multiple clusters. These clusters will most likely be adjacent to each other. As you explore, you can build a mental mapping of themes across the Galaxy.

As you explore the Galaxy using Gist, you may find a term that is unexpected or that you don't know. Or you may want to see how multiple terms are used together in documents.

- Investigate a term (or terms) of interest by highlighting it and clicking on **Query Term**. When multiple terms are highlighted, you can choose to find documents with **All terms** (AND) or **Any terms** (OR). The Gist Tool will update to show the key themes of all those documents that contain the term of interest. Look at the Galaxy to see whether the terms are scattered across the set or localized to a particular area. This will indicate how strongly this term influenced clustering.
- To see the whole text of your search results, open **Record Viewer** after querying. The upper panel shows a configurable set of columns for each document. The lower panel shows the entire content of the document highlighted in the upper panel. Search terms will be highlighted in green for easy locating.

See other guides and videos for more discussion of exploring themes in a data set.