
Using your Knowledge Base Data to Deliver Better Experiences

Kedar Joyner | UX/UI Designer | July 28, 2021



Table of Contents

01 Intro

02 Definitions

03 Page View Data

04 Search Log Data

UX/UI Designer

Kedar Joyner

UX/UI Designer with 6+ years of experience in the tech industry. I specialize in interface design within the Division of Information Technology where I help design enterprise applications for campus.



Definitions

Page View Data

Shows the number of time each individual document within a KB Group has been viewed over a given time period.

- Log into your KB Group
- Navigate to the “Stats” tab
- Click on “Doc Views by Doc” in left hand menu

The screenshot shows the 'KB Admin Tools' interface. The 'Stats' tab is active, displaying 'Internal and External Document Views for 2021'. The page includes a search filter section and a table with columns for document ID, title, creation/updated dates, and monthly view counts from January to December, plus a total. The table lists various documents such as 'KB User's Guide - Documents Tab - Status', 'KB User's Guide - Documents Tab - Show/Hide Toggles', and 'About the KnowledgeBase'.

ID	Created	Updated	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
5234	2007-01-02 19:00:00	2021-06-01 13:19:26	3,470	2,507	3,166	1,003	1,147	2,027	1,162	0	0	0	0	0	14,664
19085	2011-07-10 19:00:00	2021-06-01 13:19:26	2,800	1,978	2,445	2,273	2,127	1,405	335	0	0	0	0	0	13,363
29902	2013-04-29 08:41:31	2021-06-17 13:52:53	1,873	1,654	1,911	3,069	2,612	1,147	0	0	0	0	0	0	12,266
5238	2007-01-03 19:00:00	2021-06-19 14:22:41	3,140	2,203	2,787	631	712	1,649	800	0	0	0	0	0	11,922
3	2000-03-25 19:00:00	2021-06-01 13:18:39	1,164	1,040	1,162	1,261	2,057	1,365	1,335	0	0	0	0	0	9,384
14926	2010-08-24 19:00:00	2021-06-01 13:19:26	1,207	1,069	1,437	1,348	1,711	1,368	1,215	0	0	0	0	0	9,355
36837	2014-01-26 17:17:30	2021-06-01 13:21:00	347	304	409	528	477	4,839	399	0	0	0	0	0	7,303
27019	2013-10-24 10:14:18	2021-07-14 11:07:19	653	627	752	747	902	2,518	879	0	0	0	0	0	7,078
14927	2010-08-24 19:00:00	2021-06-19 14:15:01	909	781	1,092	1,022	1,282	1,059	899	0	0	0	0	0	7,038
60736	2016-02-12 12:36:51	2021-06-17 13:08:09	679	597	724	762	848	2,502	875	0	0	0	0	0	6,887
14767	2010-08-04 19:00:00	2021-06-01 13:19:26	842	753	985	863	973	912	827	0	0	0	0	0	6,155
52052	2015-08-08 12:24:59	2021-06-19 10:54:14	674	675	821	925	967	906	801	0	0	0	0	0	5,669
22118	2012-01-09 12:37:06	2021-06-01 13:19:45	675	616	794	765	1,036	916	747	0	0	0	0	0	5,549
66006	2016-08-15 14:50:12	2021-06-17 13:14:32	634	674	878	711	802	730	691	0	0	0	0	0	5,120
66704	2016-09-06 11:29:44	2021-06-01 13:08:58	827	607	729	691	736	753	562	0	0	0	0	0	4,905
5212	2006-12-21 19:00:00	2021-06-01 13:19:45	592	592	703	695	847	687	664	0	0	0	0	0	4,780
5233	2007-01-02 19:00:00	2021-06-01 13:19:45	672	628	704	708	723	646	601	0	0	0	0	0	4,680

How can Page View data help me?

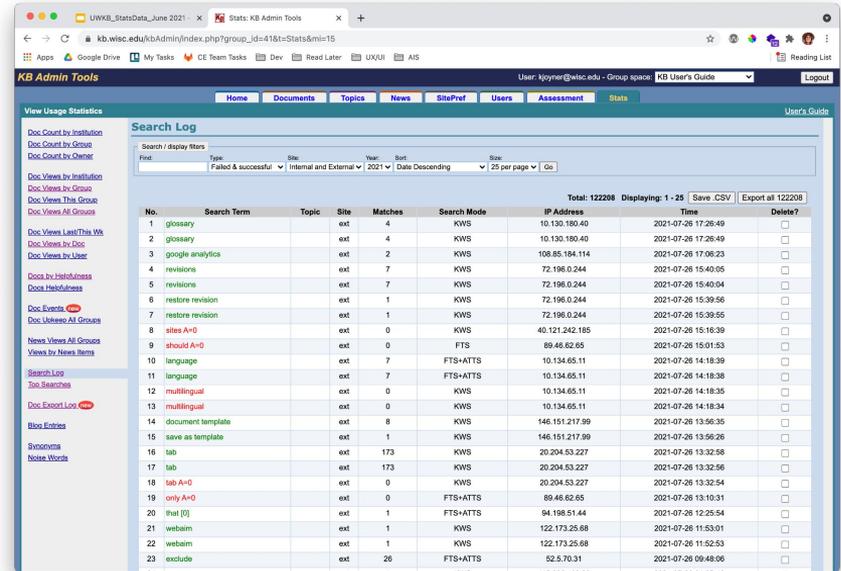
- Provides a high-level overview of audience behavior and what information they might be most interested in (i.e. informational, events, how-tos, etc.)
- Can also surface potential problems with your site. High page views do not always mean you are reaching a broad audience or that users are able to locate the information they need.
- Can be compared with search log data to understand whether people are finding answers/docs to popular searches.

Definitions

Search Log Data

Allows admins to browse through all successful (green text) and failed (red text) customer searches over a given time period.

- Log into your KB Group
- Navigate to the “Stats” tab
- Click on “Search Log” in left hand menu



The screenshot shows the KB Admin Tools interface. The top navigation bar includes tabs for Home, Documents, Topics, News, SitePref, Users, Assessment, and Stats. The left sidebar contains a menu with options like Doc Count by Institution, Doc Views by Institution, and Search Log. The main content area displays the Search Log table, which is filtered for the year 2021 and sorted by Date Descending. The table has columns for No., Search Term, Topic, Site, Matches, Search Mode, IP Address, Time, and Delete?.

No.	Search Term	Topic	Site	Matches	Search Mode	IP Address	Time	Delete?
1	glossary	ext	4	KWS	10.130.180.40	2021-07-26 17:26:49	<input type="checkbox"/>	
2	glossary	ext	4	KWS	10.130.180.40	2021-07-26 17:26:49	<input type="checkbox"/>	
3	google analytics	ext	2	KWS	108.85.194.114	2021-07-26 17:06:23	<input type="checkbox"/>	
4	revisions	ext	7	KWS	72.196.0.244	2021-07-26 15:40:05	<input type="checkbox"/>	
5	revisions	ext	7	KWS	72.196.0.244	2021-07-26 15:40:04	<input type="checkbox"/>	
6	restore revision	ext	1	KWS	72.196.0.244	2021-07-26 15:39:56	<input type="checkbox"/>	
7	restore revision	ext	1	KWS	72.196.0.244	2021-07-26 15:39:55	<input type="checkbox"/>	
8	sites A=0	ext	0	KWS	40.121.242.185	2021-07-26 15:16:39	<input type="checkbox"/>	
9	should A=0	ext	0	FTS	89.46.62.65	2021-07-26 15:01:53	<input type="checkbox"/>	
10	language	ext	7	FTS+ATTS	10.134.65.11	2021-07-26 14:18:39	<input type="checkbox"/>	
11	language	ext	7	FTS+ATTS	10.134.65.11	2021-07-26 14:18:38	<input type="checkbox"/>	
12	multilingual	ext	0	KWS	10.134.65.11	2021-07-26 14:18:35	<input type="checkbox"/>	
13	multilingual	ext	0	KWS	10.134.65.11	2021-07-26 14:18:34	<input type="checkbox"/>	
14	document template	ext	8	KWS	146.151.217.99	2021-07-26 13:56:35	<input type="checkbox"/>	
15	save as template	ext	1	KWS	146.151.217.99	2021-07-26 13:56:26	<input type="checkbox"/>	
16	tab	ext	173	KWS	20.204.53.227	2021-07-26 13:32:58	<input type="checkbox"/>	
17	tab	ext	173	KWS	20.204.53.227	2021-07-26 13:32:56	<input type="checkbox"/>	
18	tab A=0	ext	0	KWS	20.204.53.227	2021-07-26 13:32:54	<input type="checkbox"/>	
19	only A=0	ext	0	FTS+ATTS	89.46.62.65	2021-07-26 13:10:31	<input type="checkbox"/>	
20	thai [0]	ext	1	FTS+ATTS	94.198.51.44	2021-07-26 10:25:54	<input type="checkbox"/>	
21	webdm	ext	1	KWS	122.173.25.68	2021-07-26 11:53:01	<input type="checkbox"/>	
22	webdm	ext	1	KWS	122.173.25.68	2021-07-26 11:52:53	<input type="checkbox"/>	
23	exclude	ext	26	FTS+ATTS	52.5.70.31	2021-07-26 09:48:06	<input type="checkbox"/>	

How can Search Log data help me?

- Failed searches may help you identify the need to create new documents, revise existing documents, or create new/expanded keywords and synonyms.
- Additionally, failed searches can often include the most commonly misspelled words which can be added as synonyms to a document.
- Search data should help inform document titles, keywords, meta descriptions, and inform what displays on the homepage.



Page View Data

Let's learn how to export this data, strategies for analyzing the information, and steps for taking action on what you discover from your analysis.

Export Page View Data

1. Log into KB Admin
2. Choose KB Group
3. Click on “Stats” tab
4. Click on “Doc Views by Doc” from side menu
5. Increase limit to equal more than the total number of results
6. Export CSVs from the last 2 years (minimum 1 year)



Step 1

Combine both CSV files into one ring that rules them all

T17	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Title	ID	Created	Updated	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2	KB User's Gu	5234	1/2/07 19:00	6/1/21 13:19	2,987	5,199	6,136	1,882	1,085	1,243	1,077	3,605	4,626	1,241	3,764	3,267	36,112
3	KB User Gro	29902	4/29/13 9:41	6/17/21 13:52	2,251	2,356	2,875	3,019	3,816	2,889	4,036	2,080	2,895	2,108	1,689	1,876	31,890
4	KB User's Gu	5238	1/3/07 19:00	6/19/21 14:22	2,612	4,808	5,737	1,427	676	759	708	3,233	4,176	801	3,375	2,933	31,245
5	About the Kn	3	3/25/00 19:00	6/1/21 13:18	2,130	1,934	1,939	2,038	2,034	2,153	1,750	1,394	1,514	1,412	1,929	1,144	21,371
6	KB User's Gu	19085	7/10/11 19:00	6/1/21 13:19	2,963	396	419	436	381	403	380	350	3,103	3,479	3,138	2,668	18,116
7	KB User's Gu	14926	8/24/10 19:00	6/1/21 13:19	1,310	1,302	1,320	1,530	1,461	1,734	1,533	1,366	1,629	1,531	1,323	1,073	17,112
8	KB User's Gu	14797	8/4/10 19:00	6/1/21 13:19	974	1,078	1,349	1,660	1,591	1,427	1,176	1,014	1,229	1,122	975	696	14,291
9	KB User's Gu	14927	8/24/10 19:00	6/19/21 14:15	967	980	1,000	1,165	1,123	1,286	1,142	1,039	1,230	1,145	1,011	805	12,893
10	KB User's Gu	22118	1/9/12 12:37	6/1/21 13:19	980	954	911	1,026	961	1,215	1,153	1,114	1,075	1,174	913	633	12,109
11	Sample Doc:	52052	6/8/15 12:24	6/19/21 10:54	749	805	767	890	893	867	1,010	989	1,011	993	807	612	10,393
12	KB User's Gu	60736	2/12/16 12:36	6/17/21 13:08	810	775	742	847	881	948	1,023	986	929	909	781	561	10,192
13	KB User's Gu	27019	10/24/12 10:14	7/14/21 11:57	772	774	750	919	848	912	929	786	841	963	817	571	9,882
14	KB User's Gu	24488	5/25/12 16:07	6/1/21 13:19	720	728	758	833	767	853	872	802	866	885	742	503	9,329
15	KB User's Gu	5212	12/21/06 19:00	6/1/21 13:19	705	720	686	778	715	1,005	884	791	857	862	723	570	9,296
16	KB User's Gu	36926	1/29/14 9:38	6/1/21 13:19	879	3,404	1,704	436	306	320	392	417	381	429	327	213	9,208
17	KB User's Gu	5304	1/18/07 19:00	6/1/21 13:19	785	770	916	857	675	829	830	698	767	751	614	559	9,051
18	KB User's Gu	66704	9/6/16 11:29	6/1/21 13:18	692	488	690	723	791	917	703	783	864	829	884	599	8,963
19	KB User's Gu	5233	1/2/07 19:00	6/1/21 13:19	627	792	878	779	710	829	723	745	702	618	520	8,755	
20	KB Author Tr	23433	3/26/12 8:31	6/1/21 13:18	549	611	950	1,111	1,125	854	706	599	744	608	576	381	8,714
21	KB User's Gu	5722	4/10/07 19:00	6/1/21 13:19	616	742	628	809	597	792	680	664	746	735	581	407	7,997
22	KB User's Gu	36842	1/28/14 17:17	6/1/21 13:19	561	536	516	567	666	637	656	724	732	773	644	463	7,475
23	KB User's Gu	60996	8/15/16 14:50	6/17/21 13:14	919	522	488	517	436	613	661	691	744	739	613	510	7,453
24	KB User's Gu	5235	1/2/07 19:00	5/17/21 16:12	664	594	634	767	564	673	671	617	670	657	357	431	7,399
25	KB Author Tr	23277	3/16/12 13:55	5/17/21 16:11	625	608	612	783	623	748	659	467	663	609	472	419	7,288
26	KB User's Gu	4643	4/27/06 19:00	6/1/21 13:19	590	683	688	702	514	603	662	640	645	616	466	389	7,198
27	KB User's Gu	5344	1/29/07 19:00	6/1/21 13:19	585	589	538	597	577	724	682	651	648	705	513	350	7,159
28	KB User's Gu	5338	1/29/07 19:00	6/1/21 13:19	547	533	556	490	589	756	712	671	627	663	527	403	7,074
29	KB User's Gu	23534	3/30/12 8:49	6/19/21 14:10	487	526	531	642	579	650	583	564	689	582	572	510	6,915
30	Poetry - Robi	36746	1/28/14 17:02	6/1/21 13:21	541	461	621	712	584	575	573	515	545	569	431	384	6,511
31	KB User's Gu	69832	1/9/17 14:27	6/17/21 15:06	477	475	447	486	562	574	639	624	552	715	546	304	6,401
32	KB Users Gu	13301	2/10/10 19:00	5/17/21 16:14	578	524	543	576	501	607	614	546	572	548	463	323	6,395
33	KB Author Tr	23140	3/14/12 8:15	6/1/21 13:18	511	489	487	582	535	651	554	442	564	611	550	414	6,390
34	KB User's Gu	8531	11/17/08 19:00	6/1/21 13:19	511	538	524	561	542	572	527	482	649	557	461	440	6,359
35	KB User's Gu	14925	8/24/10 19:00	6/1/21 13:19	450	518	439	506	519	522	635	595	600	583	484	378	6,229
36	KB User's Gu	5307	1/18/07 19:00	6/1/21 15:07	528	531	578	579	521	500	586	549	506	543	459	328	6,208
37	KB User's Gu	3469	2/8/05 19:00	6/1/21 13:19	450	456	478	458	491	503	627	618	607	790	412	275	6,165
38	KB User's Gu	15019	8/30/10 19:00	5/17/21 16:12	576	503	527	581	484	581	441	445	526	579	451	313	6,007

Step 2

Delete all columns except for Title and Total

	Title	Total
2	KB User's Guide - Documents Tab - Status	36,112
3	KB User Group Meeting	31,890
4	KB User's Guide - Documents Tab - Overview	31,245
5	About the KnowledgeBase	21,371
6	KB User's Guide - Documents Tab - Show/H	18,116
7	KB User's Guide - Documents Tab - ImportH	17,112
8	KB User's Guide - Documents Tab - Guide to	14,291
9	KB User's Guide - Documents Tab - Guide to	12,893
10	KB User's Guide - Topics - The Batch Topic A	12,109
11	Sample Doc: IncludeDoc - Violin Parts	10,393
12	KB User's Guide - SitePref Tab - Content Mo	10,192
13	KB User's Guide - SitePref Tab - Side Modul	9,882
14	KB User's Guide - SitePref Tab- Quick Search	9,329
15	KB User's Guide - Topics Tab - Enable the To	9,296
16	KB User's Guide - Documents Tab - Referen	9,208
17	KB User's Guide - Documents Tab - Docume	9,051
18	KB User's Guide - Advanced HTML - Decisio	8,963
19	KB User's Guide - General Info - Table of Co	8,755
20	KB Author Training - Body Field - Design (W	8,714
21	KB User's Guide - Documents Tab - Uploadi	7,997
22	KB User's Guide - Documents Tab - Review	7,475

Step 3

Select a cell where you want the merged data to appear

A	B	C
Title	Total	
KB User's Guide - Documents Tab - Status	36,112	
KB User Group Meeting	31,890	
KB User's Guide - Documents Tab - Overview	31,245	
About the KnowledgeBase	21,371	
KB User's Guide - Documents Tab - Show/H	18,116	
KB User's Guide - Documents Tab - ImportH	17,112	
KB User's Guide - Documents Tab - Guide to	14,291	
KB User's Guide - Documents Tab - Guide to	12,893	
KB User's Guide - Topics - The Batch Topic	12,109	

Step 4

Go to Data > Consolidate and select "Sum" as the function, check "Top Row", and select columns A and B

Consolidate

Function: Sum

Reference: \$A:\$B

All references:

Select...

Create links to source data

Use labels in:

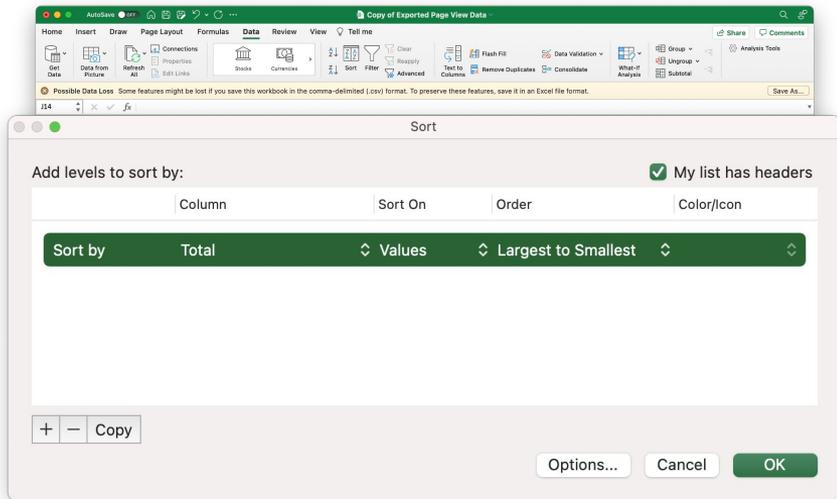
Top row

Left column

Close OK

Step 5

Copy the output into a new tab and sort the data from smallest to largest in the Total column



Done! 🎉

Now you're seeing which KB Documents have the highest page views over a 2 year time period. You're ready to analyze!

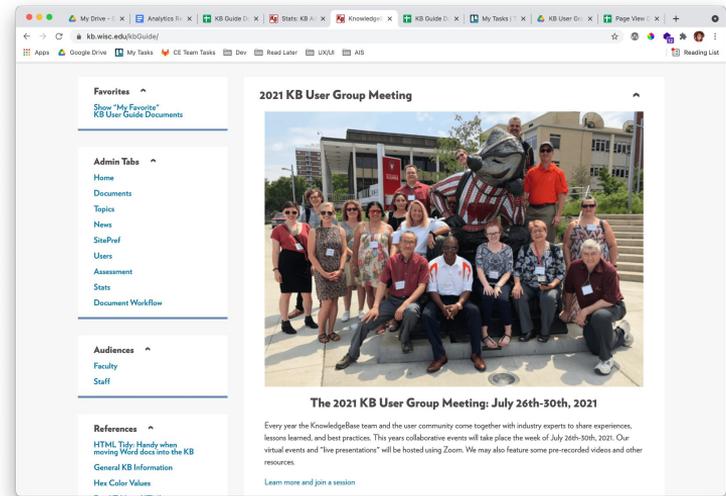
Content Modules

Ignore any page views that are given to Content Module documents. Every time a user visits a page with a content module, it's recorded as a page view for that particular document.

For documents with high page views, check if they are content module first by going to **KB Admin > Site Prefs > Content modules**

Example

KB User Group Meeting has 48,653 page views, but it's a Content Module so that number is inflated.



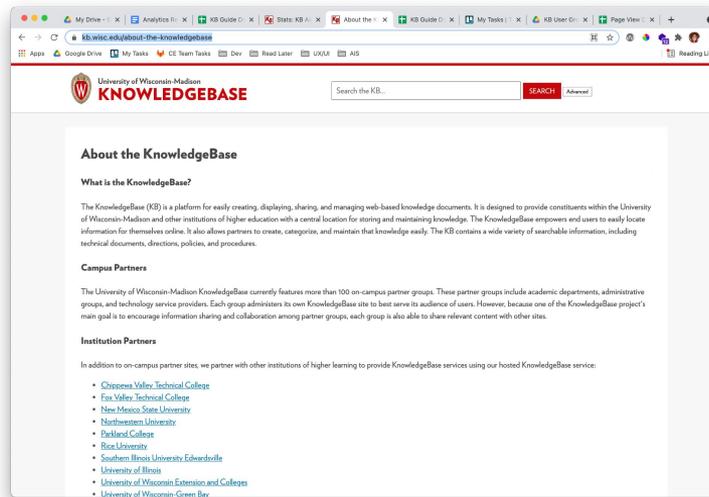
“Linked From” Documents

Documents that are linked from other documents may have inflated page views because they encourage more traffic. However, this does not mean people find the information on these documents useful.

If your document is being referenced in other documents, the option to view “Linked from” will appear at the bottom of the doc in the admin.

Example

This document has 30,764 views and is linked from 74 other docs, but it has 591 helpful votes and 437 unhelpful





Analyze the Data

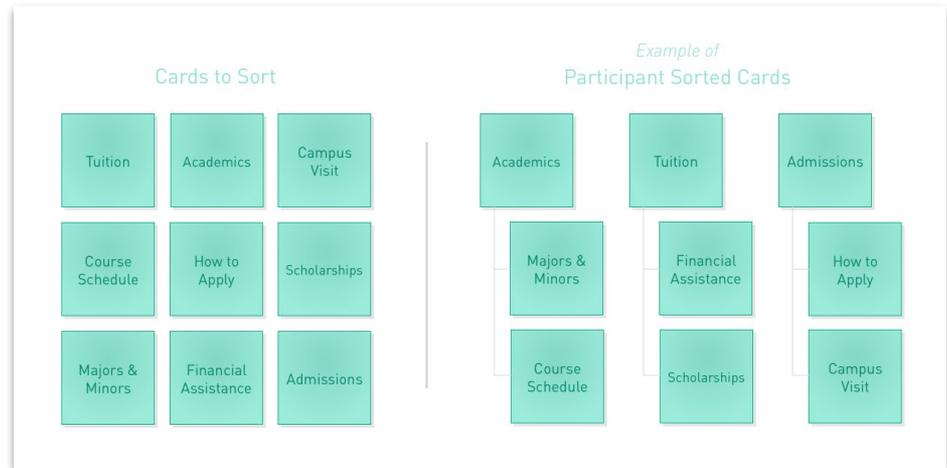
Start small and look at the top 25-50 most viewed documents.

Questions to Ask	Notes
Does anything look odd or unexpected?	Look for things that make you think, "Oh, interesting!" to investigate further.
Does your KB homepage reflect the documents that people are viewing the most?	Helpful to compare this data with your existing homepage to see if popular documents are missing from view or need to be placed higher on the page.
Are these documents easy to find? And is there anything confusing about them?	If they're not referenced on the homepage, should they be? Are they easily searchable? Do they contain relevant keywords and synonyms to improve searchability?
Can these documents be grouped together in logical categories (if not already) that might help users find the information they need more quickly?	You may find that based on what you find, your homepage could be improved by improving navigation and/or groupings of content under categories that make sense to your audience. Consider a card sort exercise to determine these categories.

Card Sorting

Card sorting is a method used to help design or evaluate the information architecture of a site.

In a card sorting session, participants organize topics into categories that make sense to them and they may also help you label these groups.



Example

For the KB User's Guide we noticed based on the top page views, that content on the homepage was organized from least important to most important, and we want to flip this order.

The first 3 Top Docs are also content modules, so they likely aren't top docs in the truest sense.

Sidebar content also does not contain the themes we found in the top page views and searches.

The screenshot displays a KnowledgeBase interface with a sidebar on the left and a main content area on the right. The sidebar contains several sections: Favorites (linking to 'My Favorites' and 'KB User Guide Documents'), Admin Tabs (Home, Documents, Topics, News, SitePref, Users, Assessment, Stats, Document Workflow), Audiences (Faculty, Staff), References (HTML Tab Handy when meeting, How to add docs into the KB, General KB Information, Hex Color Values, Excel Table to HTML, HTML Primer), Helpful KB Tips (User's Guide Keyword Index, LifeCycle of your KB Document, WYSIWYG Editor Controls, HTML Editor Quick Code Buttons), KB Trivia Boards (Known Issues, Feature Requests, Product Roadmap), and 2021 KB Cafe (UW-Madison Only) with a 'Join Us!' link.

The main content area features a large photo of a group of people at a '2021 KB User Group Meeting'. Below the photo is a caption: 'The 2021 KB User Group Meeting: July 26th-30th, 2021'. A short paragraph follows, stating: 'Every year the KnowledgeBase team and the user community come together with industry experts to share experiences, lessons learned, and best practices. This year collaborative events will take place the week of July 26th-30th, 2021. Our virtual events and "Pre presentation" will be hosted using Zoom. We may also feature some pre-recorded videos and other resources.' A link 'Learn more and join a session' is provided.

Below the photo is a 'Documents Recently Updated' section with a list of 8 items, including 'KB User's Guide - Documents Tab - Left Navigation Bar Links', '2021 KnowledgeBase (KB) User Group Meeting', 'KB User Group Meeting', 'KB User's Guide - Documents Tab - Import & Export', 'KB User's Guide - Stats Tab - Doc Events', 'KB User's Guide - Stats Tab - Overview of Links', 'KB User's Guide - UW Theme for KB: Features', and 'KB User's Guide - SitePref Tab - Side Modules and Side Module Links'. A 'MORE RECENT DOCS' button is at the bottom right.

The 'Top Documents' section lists 5 items: 'KB User Group Meeting', '2021 KnowledgeBase (KB) User Group Meeting', 'KB User's Guide - Documents Tab - ImportHTML button', 'About the KnowledgeBase', and 'Sample Doc: IncludeDoc - Visuln Parts'. A 'MORE TOP DOCS' button is at the bottom right.

The 'KB Basics - Start here!' section lists 4 items: 'KB Author Training - Overview', 'KB Author Training - Creating a New Document', 'KB Author Training - Attachments and Revisions', and 'My KB Group Space: A First Look'.

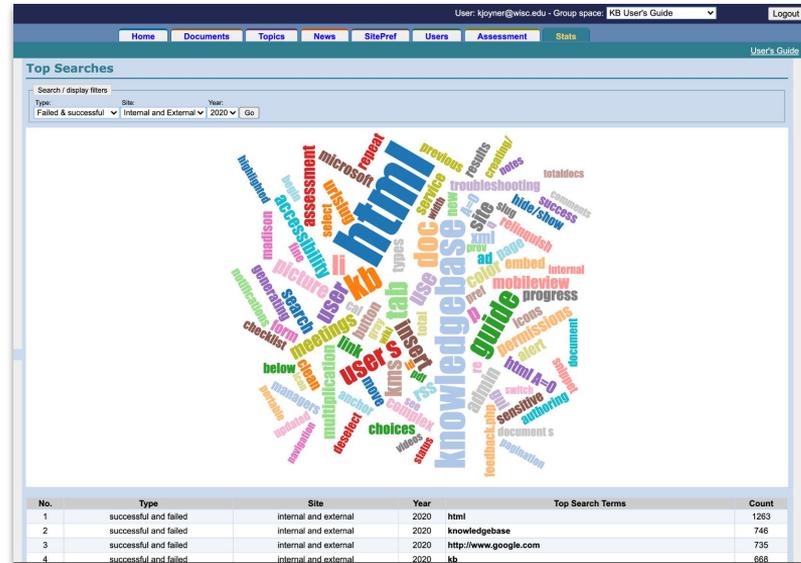


Search Log Data

Let's learn how to export Search Log data and filter out pesky bot traffic to have a more accurate picture of what your users are searching for on your KB site.

Why don't we just view the "Top Searches" log?

- Can only view one year of data at a time
- Doesn't filter out bot traffic. Discovered that roughly half of searches originated from non-human traffic or untrustworthy IP addresses
- Still provides a useful picture, but not as accurate



Export Search Log Data

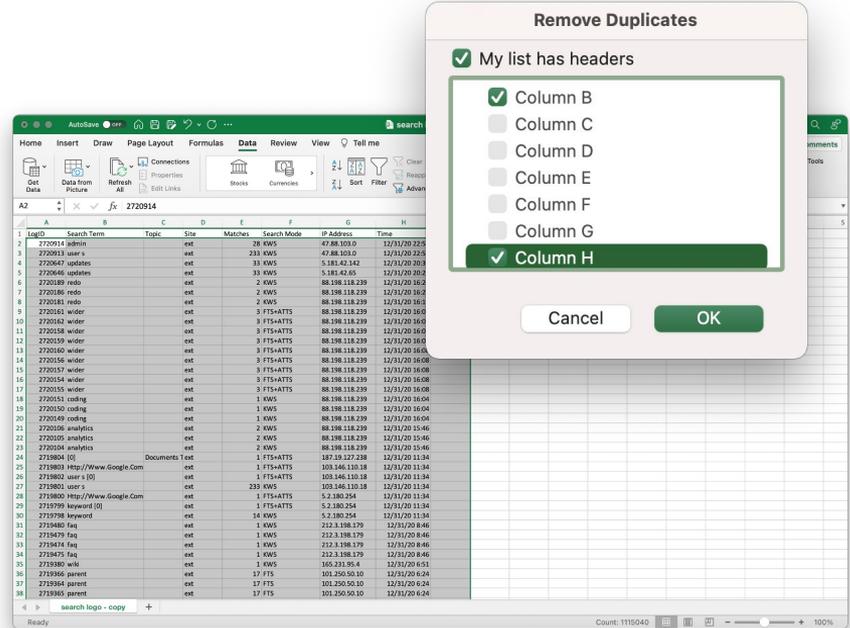
1. Log into KB Admin
2. Choose KB Group
3. Click on “Stats” tab
4. Click on “Search Log” from side menu
5. Filter log data by “Successful” or “Failed” respectively
6. Click on the “Export All” button
7. Export CSVs from the last 2 years (minimum 1 year)
8. Combine both files into one

Remove Duplicates

Step 1

Often bots will make the same search multiple times at once or within seconds/minutes of each other.

So we'll remove duplicate entries that contain the search term and occur at the same time.



The screenshot shows an Excel spreadsheet with a 'Remove Duplicates' dialog box open. The dialog box has a title 'Remove Duplicates' and a checked option 'My list has headers'. Below this, there is a list of columns with checkboxes: Column B (checked), Column C, Column D, Column E, Column F, Column G, and Column H (checked). At the bottom of the dialog are 'Cancel' and 'OK' buttons. The spreadsheet data is as follows:

ID	Search Term	Topic	Site	Matches	Search Month	IP Address	Time
2720914	admin	ent	28 KWS	47.88.103.0	12/31/20 22:25		
2720913	user s	ent	233 KWS	47.88.103.0	12/31/20 22:25		
2720943	updates	ent	33 KWS	5.181.42.142	12/31/20 20:19		
2720646	updates	ent	33 KWS	5.181.42.65	12/31/20 20:22		
2720589	redo	ent	2 KWS	88.198.118.239	12/31/20 16:02		
2720586	redo	ent	2 KWS	88.198.118.239	12/31/20 16:02		
2720585	redo	ent	2 KWS	88.198.118.239	12/31/20 16:02		
2720583	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:02		
2720582	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:02		
2720580	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:02		
2720559	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:02		
2720560	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:02		
2720566	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:08		
2720552	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:08		
2720554	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:08		
2720555	wider	ent	3 FTS+ATTS	88.198.118.239	12/31/20 16:08		
2720553	coding	ent	1 KWS	88.198.118.239	12/31/20 16:04		
2720500	coding	ent	1 KWS	88.198.118.239	12/31/20 16:04		
2720505	coding	ent	1 KWS	88.198.118.239	12/31/20 16:04		
2720506	analytics	ent	2 KWS	88.198.118.239	12/31/20 15:46		
2720505	analytics	ent	2 KWS	88.198.118.239	12/31/20 15:46		
2720504	analytics	ent	2 KWS	88.198.118.239	12/31/20 15:46		
2719860 (0)		Documents Text	1 FTS+ATTS	103.146.110.138	12/31/20 11:34		
2719863	http://www.google.com	ent	1 FTS+ATTS	103.146.110.138	12/31/20 11:34		
2719860	user s (0)	ent	1 FTS+ATTS	103.146.110.138	12/31/20 11:34		
2719860	user s	ent	233 KWS	103.146.110.138	12/31/20 11:34		
2719860	http://www.google.com	ent	1 FTS+ATTS	5.2.180.254	12/31/20 11:34		
2719799	keyword (0)	ent	1 FTS+ATTS	5.2.180.254	12/31/20 11:34		
2719798	keyword	ent	14 KWS	5.2.180.254	12/31/20 11:34		
2719480	faq	ent	1 KWS	212.3.198.179	12/31/20 8:46		
2719479	faq	ent	1 KWS	212.3.198.179	12/31/20 8:46		
2719478	faq	ent	2 KWS	212.3.198.179	12/31/20 8:46		
2719475	faq	ent	1 KWS	212.3.198.179	12/31/20 8:46		
2719380	whi	ent	1 KWS	165.233.95.4	12/31/20 6:51		
2719366	parent	ent	17 FTS	101.205.50.10	12/31/20 6:24		
2719366	parent	ent	17 FTS	101.205.50.10	12/31/20 6:24		
2719360	parent	ent	17 FTS	101.205.50.10	12/31/20 6:24		

Note: It's also helpful at this time to remove cells that are blank or contain Excel errors.

Remove Special Characters

Step 2

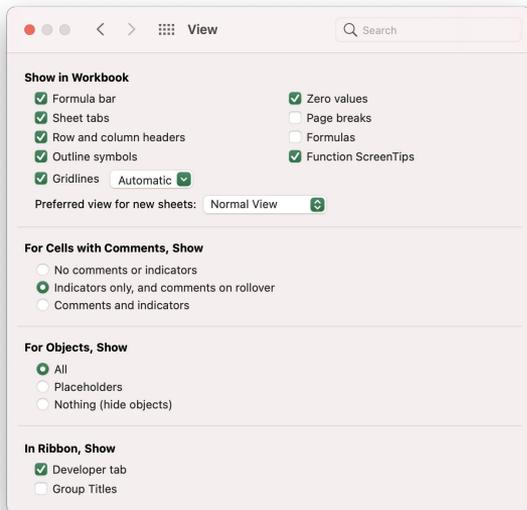
Bots will often make SQL injections (e.g. "A=0") and special characters (e.g. "/", "[0]", "!+{ }) into their searches.

In the upcoming steps, we'll remove all searches that contain special characters to help filter out bot traffic and clean the data as best we can.

SPECIAL CHARACTERS													
a A	b B	c C	d D	e E	f F	g G	h H	i I	j J	k K	l L	m M	n N
o O	p P	q Q	r R	s S	t T	u U	v V	w W	x X	y Y	z Z	A Shift+A	B Shift+B
C Shift+C	D Shift+D	E Shift+E	F Shift+F	G Shift+G	H Shift+H	I Shift+I	J Shift+J	K Shift+K	L Shift+L	M Shift+M	N Shift+N	O Shift+O	P Shift+P
Q Shift+Q	R Shift+R	S Shift+S	T Shift+T	U Shift+U	V Shift+V	W Shift+W	X Shift+X	Y Shift+Y	Z Shift+Z	1 1	2 2	3 3	4 4
5 5	6 6	7 7	8 8	9 9	0 0	! Shift+1	@ Shift+2	# Shift+3	\$ Shift+4	% Shift+5	^ Shift+6	& Shift+7	* Shift+8
(Shift+9) Shift+0	, ,	. .	; ;	: Shift+;	- -	/ /	? Shift+/ ?	' '	" Shift+' "	~ Alt+0161	¡ Alt+0191	— Alt+0173
+ Shift+=	× Alt+0215	÷ Alt+0247	= =	± Alt+0177	< Shift+<	> Shift+>	[] 	{ Shift+ {	}	' Alt+0145	' Alt+0146	“ Alt+0147

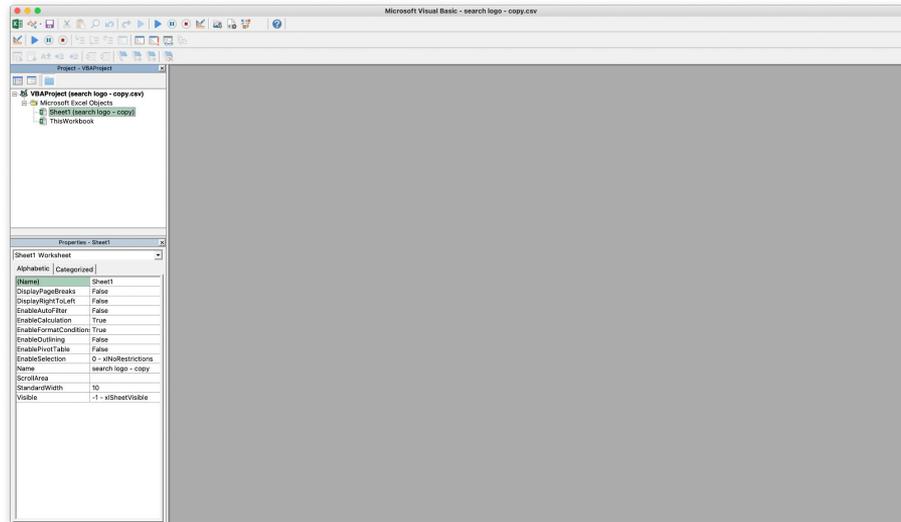
Step 3

In your Excel file click on “Excel” in the top menu > Preferences > View and check “Developer tab”



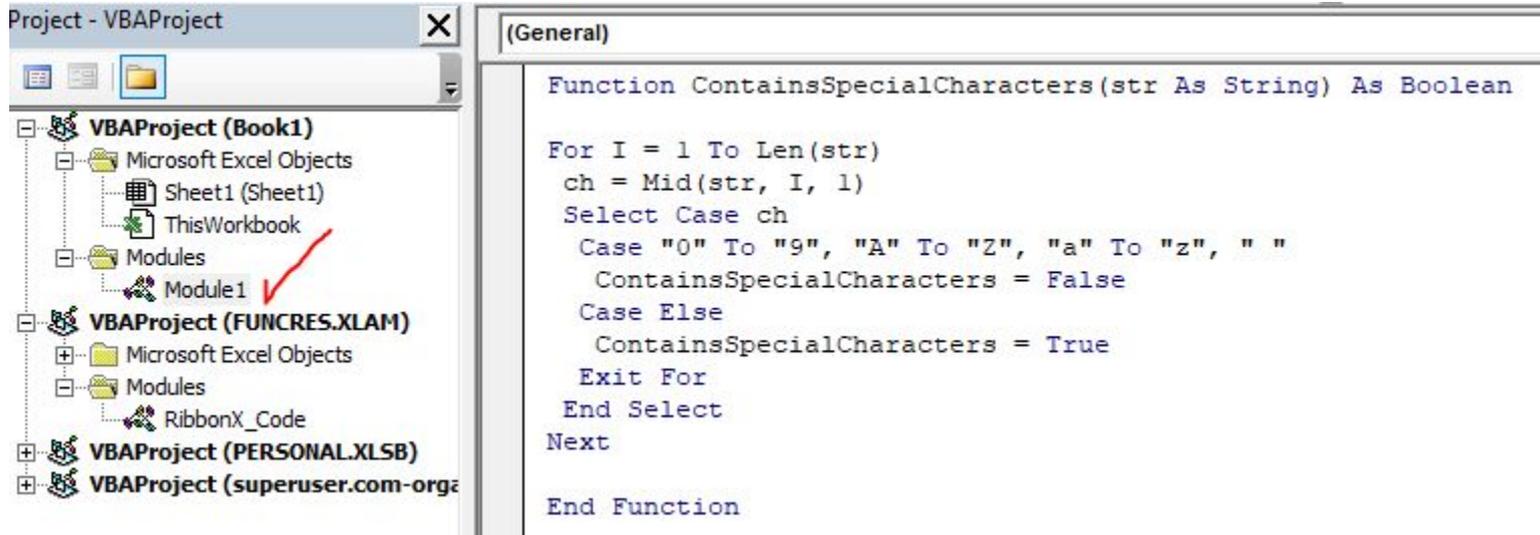
Step 4

Now click on the “Developer” tab > Visual Basic > and then you will see the VB editor



Step 5

Insert a module from the Insert menu. Copy the code below and paste it into the module. Click save, and now your function is ready to be used!



The screenshot displays the VBA Project Explorer on the left and the Code Editor on the right. The Project Explorer shows a tree view with 'VBAProject (Book1)' expanded to show 'Microsoft Excel Objects' (containing 'Sheet1 (Sheet1)' and 'ThisWorkbook') and 'Modules' (containing 'Module1', which is highlighted with a red checkmark). Below it are other projects: 'VBAProject (FUNCRES.XLAM)', 'VBAProject (PERSONAL.XLSB)', and 'VBAProject (superuser.com-org)'. The Code Editor shows the following VBA code:

```
(General)  
  
Function ContainsSpecialCharacters(str As String) As Boolean  
  
For I = 1 To Len(str)  
    ch = Mid(str, I, 1)  
    Select Case ch  
        Case "0" To "9", "A" To "Z", "a" To "z", " "  
            ContainsSpecialCharacters = False  
        Case Else  
            ContainsSpecialCharacters = True  
    End Select  
    Exit For  
End Select  
Next  
  
End Function
```

Step 6

Insert a new column in your primary workbook next to the “Search Term” column

LogID	Search Term	Special Characters?	Topic	Site	Matches	Search Moc
2720914	admin			ext	28	KWS
2720913	user s			ext	233	KWS
2720647	updates			ext	33	KWS
2720646	updates			ext	33	KWS
2720189	redo			ext	2	KWS
2720186	redo			ext	2	KWS
2720181	redo			ext	2	KWS
2720161	wider			ext	3	FTS+ATTS
2720158	wider			ext	3	FTS+ATTS
2720156	wider			ext	3	FTS+ATTS
2720154	wider			ext	3	FTS+ATTS
2720151	coding			ext	1	KWS
2720150	coding			ext	1	KWS
2720149	coding			ext	1	KWS
2720106	analytics			ext	2	KWS
2720105	analytics			ext	2	KWS
2720104	analytics			ext	2	KWS
2719804	[0]		Documents	ext	1	FTS+ATTS
2719803	Http://Www.Google.Com			ext	1	FTS+ATTS
2719802	user s [0]			ext	1	FTS+ATTS
2719801	user s			ext	233	KWS
2719800	Http://Www.Google.Com			ext	1	FTS+ATTS
2719799	keyword [0]			ext	1	FTS+ATTS
2719798	keyword			ext	14	KWS
2719480	faq			ext	1	KWS
2719479	faq			ext	1	KWS
2719474	faq			ext	1	KWS
2719380	wiki			ext	1	KWS
2719366	parent			ext	17	FTS
2719364	parent			ext	17	FTS
2719363	parent			ext	17	FTS
2719362	parent			ext	17	FTS
2719361	parent			ext	17	FTS
2719360	parent			ext	17	FTS

Step 7

Write the below in Cell 2 in the new column you just created.

LogID	Search Term	Special Characters?	Topic
2720914	admin	=ContainsSpecialCharacters(B2)	
2720913	user s	FALSE	
2720647	updates	FALSE	
2720646	updates	FALSE	
2720189	redo	FALSE	
2720186	redo	FALSE	
2720181	redo	FALSE	
2720161	wider	FALSE	
2720158	wider	FALSE	
2720156	wider	FALSE	

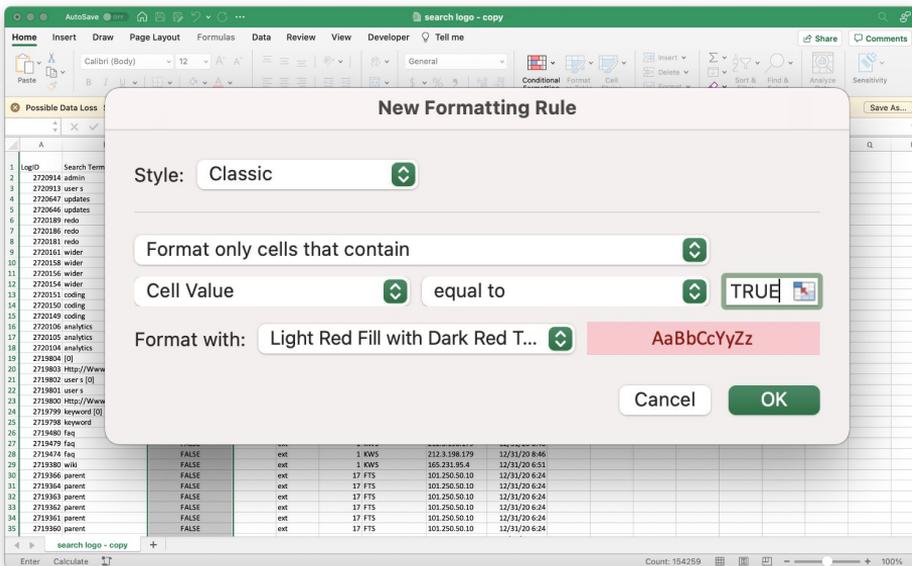
Step 8

Double click the bottom right corner of Cell 2 to apply the function to the entire row. It will display TRUE for strings that contain a special character and FALSE for those that do not.

A	B	C
LogID	Search Term	Special Characters?
2720914	admin	FALSE
2720913	user s	FALSE
2720647	updates	FALSE
2720646	updates	FALSE
2720189	redo	FALSE
2720186	redo	FALSE
2720181	redo	FALSE
2720161	wider	FALSE
2720158	wider	FALSE
2720156	wider	FALSE
2720154	wider	FALSE
2720151	coding	FALSE
2720150	coding	FALSE
2720149	coding	FALSE
2720106	analytics	FALSE
2720105	analytics	FALSE
2720104	analytics	FALSE
2719801	user s	FALSE
2719798	keyword	FALSE
2719480	faq	FALSE
2719479	faq	FALSE
2719474	faq	FALSE
2719380	wiki	FALSE
2719366	parent	FALSE
2719364	parent	FALSE
2719363	parent	FALSE
2719362	parent	FALSE
2719361	parent	FALSE

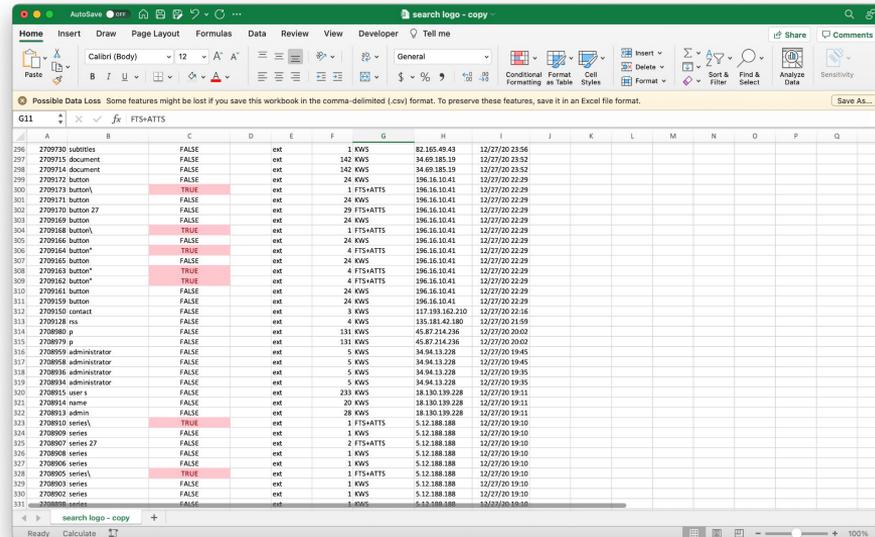
Step 9

Use the “Conditional Formatting” feature to highlight a cell red if it’s equal to the word “TRUE”



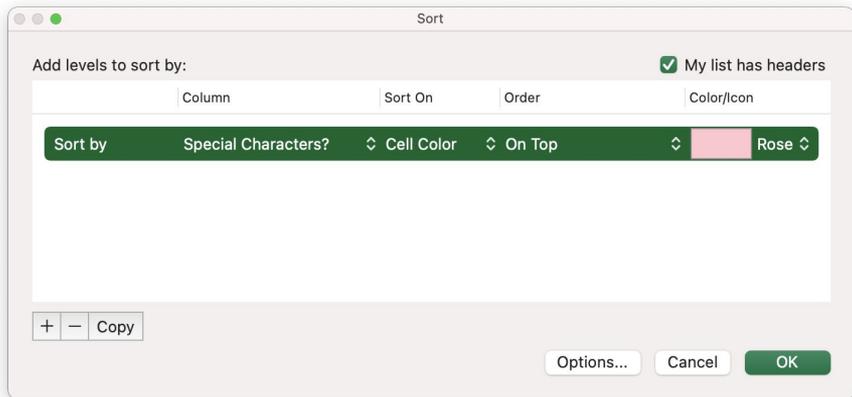
Step 10

All cells that contain “TRUE” should now be highlighted red.



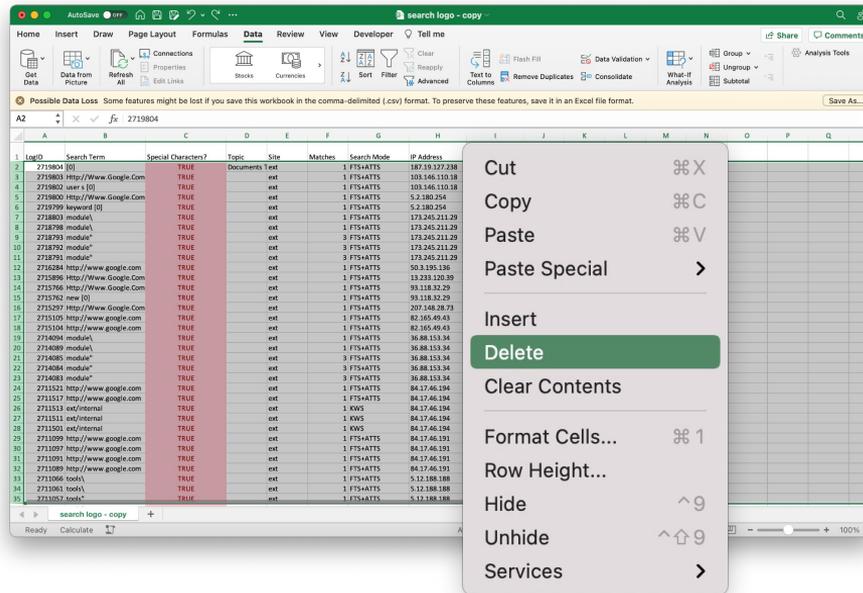
Step 11

Select your entire spreadsheet and sort your spreadsheet based on the highlighted cells.



Step 12

Select all of the cells highlighted red and delete.



10,404

I got rid of 10,404 rows after cleaning my Search Log data. If you were to check every IP address, even more would be removed but that's not an efficient use of time with thousands of results. With this method, we *at least* have a more accurate data set than we did before.

Next, we'll walk through how to [find all of the unique search values](#) and count how many times these words were searched.



Step 14

Because there are multiples of the same search terms, we want to create a list of each search term that contains only one of each. We'll use the "Unique" formula like shown here:

Add the formula to the first cell in the second column.

The screenshot shows an Excel spreadsheet with two columns, A and B. Column A contains a list of search terms: admin, user s, updates, redo, wider, coding, analytics, user s, keyword, faq, wiki, parent, parent, parent, parent. The formula bar at the top shows the formula =unique(A:A) being entered into cell B1.

	A	B
1	Search Term	=unique(A:A)
2	admin	
3	user s	
4	updates	
5	updates	
6	redo	
7	redo	
8	redo	
9	wider	
10	wider	
11	wider	
12	wider	
13	coding	
14	coding	
15	coding	
16	analytics	
17	analytics	
18	analytics	
19	user s	
20	keyword	
21	faq	
22	faq	
23	faq	
24	wiki	
25	parent	
26	parent	
27	parent	
28	parent	

Press enter to apply the function to the entire row.

The screenshot shows the same Excel spreadsheet as before, but now the formula has been applied to the entire row. The results are shown in column B: admin, user s, updates, redo, wider, coding, analytics, keyword, faq, wiki, parent, tableizer, displaying KB, module, conversion, Kbadmin, interest, module 27, recordings, listener, ui, tips, Access, Permissions, Excel table, edit an existir.

	A	B
1	Search Term	Search Term
2	admin	admin
3	user s	user s
4	updates	updates
5	updates	redo
6	redo	wider
7	redo	coding
8	redo	analytics
9	wider	keyword
10	wider	faq
11	wider	wiki
12	wider	parent
13	coding	tableizer
14	coding	displaying KB
15	coding	module
16	analytics	conversion
17	analytics	Kbadmin
18	analytics	interest
19	user s	module 27
20	keyword	recordings
21	faq	listener
22	faq	ui
23	faq	tips
24	wiki	Access
25	parent	Permissions
26	parent	Excel table
27	parent	edit an existir
28	parent	

Counting the Occurrences

Step 15

Now that we have a unique set of Search Terms, we want to count how many times each term occurs. Starting by creating a new column titled “Count”

	A	B	C
1	Search Term	Search Term	Count
2	admin	admin	
3	user s	user s	
4	updates	updates	
5	updates	redo	
6	redo	wider	
7	redo	coding	
8	redo	analytics	
9	wider	keyword	

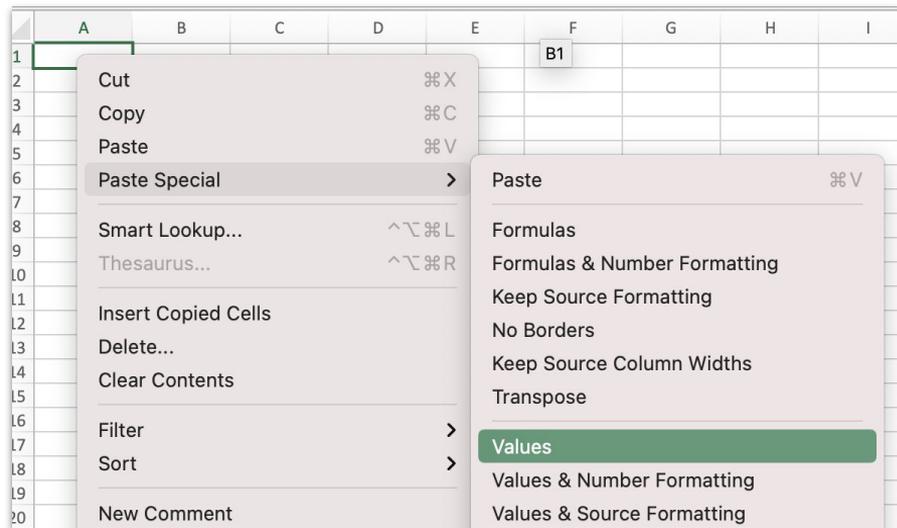
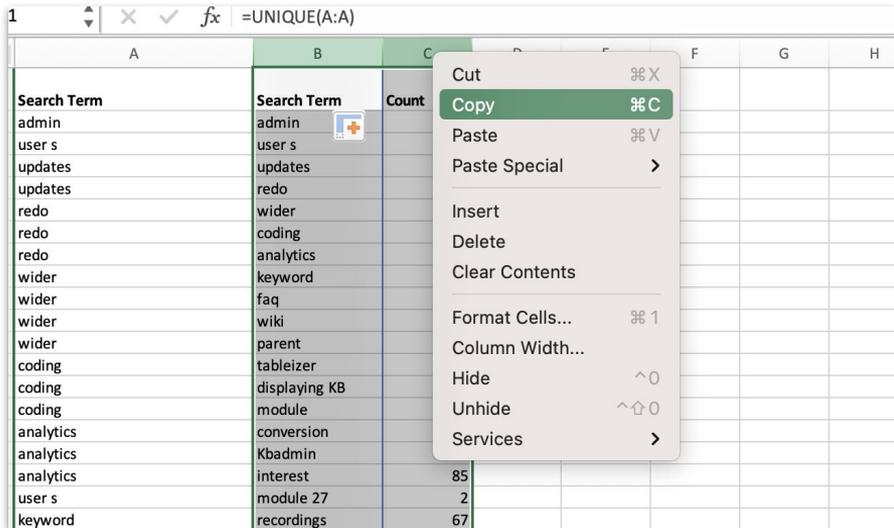
Step 16

Use Excel's “=COUNTIF” function to count how many times each unique search term occurs in Column A.

	A	B	C	D
1	Search Term	Search Term	Count	
2	admin	admin	=COUNTIF(A:A, B2)	
3	user s	user s	480	
4	updates	updates	149	
5	updates	redo	143	
6	redo	wider	62	
7	redo	coding	63	
8	redo	analytics	74	
9	wider	keyword	81	
10	wider	faq	107	
11	wider	wiki	205	
12	wider	parent	81	
13	coding	tableizer	84	
14	coding	displaying KB	1	
15	coding	module	160	
16	analytics	conversion	168	
17	analytics	Kbadmin	61	
18	analytics	interest	85	
19	user s	module 27	2	

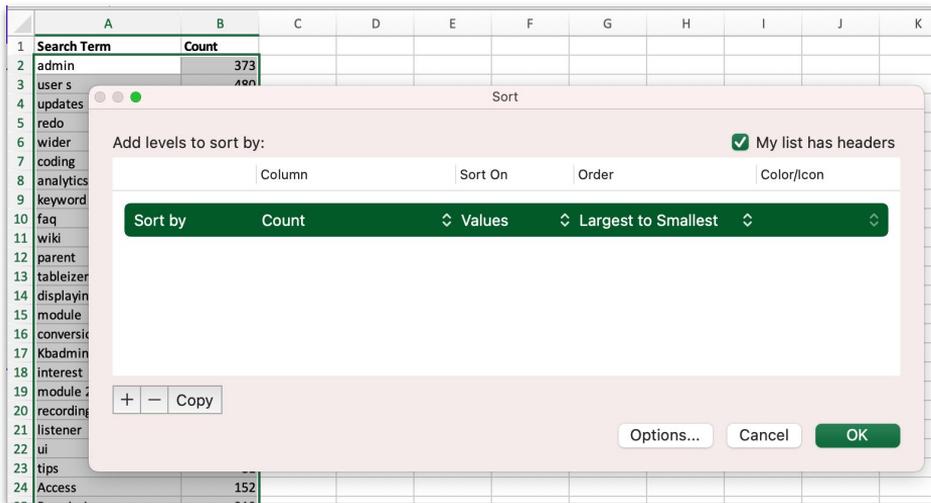
Step 16

Copy column B and C and paste into a new tab, but we want to “Paste Special” and then choose “Value”. We want static version of the dataset, otherwise the embedded formula will try to recalculate the data when we sort the sheet leading to errors.



Step 17

Sort the data based on the value of the Count column from largest to smallest.



Step 18

Review your file and remove anything that's clearly spam (i.e. "???"), still contains special characters, or empty cells. Now you have a more accurate dataset to work with!

	Search Term	Count
1	html	1218
2	knowledgebase	740
3	kb	625
4	guide	584
5	doc	582
6	li	494
7	tab	483
8	user s	480
9	insert	423
10	kms	414
11	user	413
12	use	407
13	picture	387
14	site	374
15	admin	373
16	p	364
17	meetings	364
18	accessibility	344
19	rss	338
20	search	331
21		

Analyze the Data

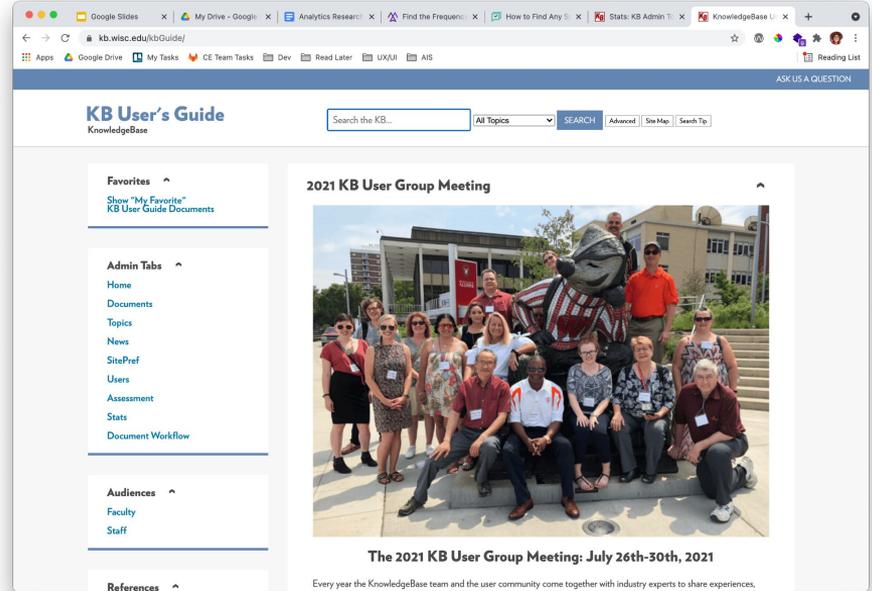
Start small and look at the top 50 - 100 searches. Look for those “Oh, interesting!” moments.

Questions to Ask	Notes
Do the top search terms also match the top documents?	If they don't, you may need to update the site to show those things in order to help users find what they're looking for most.
When looking at failed searches, what are keywords people are searching for? Do these keywords need to be added and/or do new documents need to be created?	Looking at failed searches can help you you find any keywords, synonyms, or documents that might need to be added/updated.
Also for failed searches, are there any common misspellings?	Consider adding these misspellings as synonyms to your documents(s) so that users still see related results.
Can you group these search terms into common themes? What patterns do you notice in these common groupings?	This may help you understand how users think about the content for your particular service/group, and you can use this information to organize content in a way that they related to.

Example

For the KB User's Guide we grouped searches into the following categories:

- How to design with the KB
- Understanding what things are
- Content sharing
- Notifications
- KB meeting information
- Authoring
- Training
- Batch features/migrations
- Searching for university-wide information, rather than just KB





Questions?

"A user interface is like a joke. If you have to explain it, it's not that good." - Martin LeBlanc

Contact

kedar.joyner@wisc.edu

