Are You Missing Out?
9 Ways to Test Your Search Engine

(All results from 28 February 2018)

You can be a hero in your organization by introducing the breakthrough literature deep discovery application, Qinsight. Qinsight uses artificial intelligence, coupled with visual analytics of text within articles, to give you unprecedented insights. In the pharmaceutical field, about $80 billion[1] is wasted every year rediscovering information that was previously published – reflecting the ineffectiveness of traditional search solutions.

Following are several examples comparing Qinsight to other search engines. Try the ones that are relevant to your and your organization's needs. Show these to your colleagues and managers. Be an innovation leader.

1 EXAMPLE: What genes are associated with melanoma?

Answering questions is an important aspect of effective and efficient literature discovery. Oftentimes, we should take a step back and get a general picture of the literature before diving in too deeply with looking for what we thought we should find. The bigger picture can provide insights that are otherwise missed.

The example in this section, **What genes are associated with melanoma?** is just one type of question that you should be able to answer with your search engine.

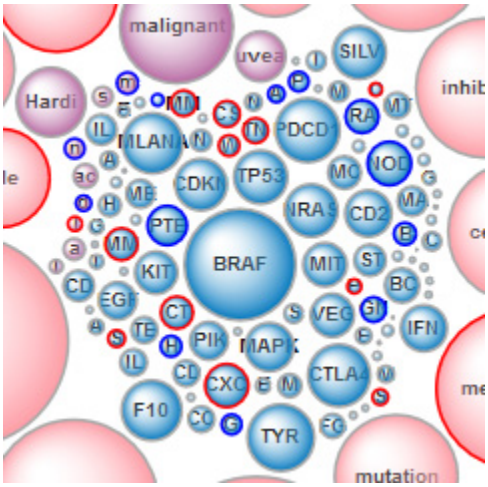
1.1 QINSIGHT

With this example, the depth of Qinsight's discovery is obvious.

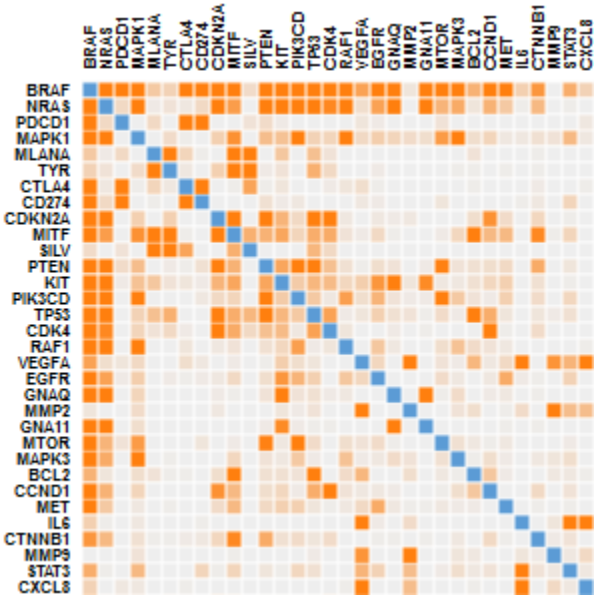
- The BioAI™ engine recognizes that by asking "What genes", the user is interested in actual genes, not the generic terms "gene" or "genetic".
- The results show actual genes and the different types of melanoma.
- It is immediately obvious why the results are relevant.
- Over 66,000 results, but easy to explore through the filters.
- Clicking on a gene or a type of melanoma refines the results.

Predictive visual analytics help you understand what the results set is all about, identify the trends, and see hidden connections. Although you get a nice list of the genes that are identified in the literature as having some relationship to melanoma, with the Concept Trends visualization (a portion of which is shown below), you can immediately see which genes are most prevalent. Sometimes more importantly, the genes with the red borders are ones that the Qinsight AI predicts will become more important in the future.

[1] Per the European Patent Office



The Concepts Connection visualization provides a deeper understanding, that is simply not possible to gain from the list of results. A portion of this visualization is shown below. The significance of the interaction between any pair of genes in contributing to melanoma is indicated by the increasing intensity of the orange color. In essence, this is a 2D representation of the gene network involved. The underlying data can be exported for use in other analytic tools.



The usefulness of these visual analytics approach is several-fold. You can get a global picture of the role of genes in melanoma, and then drill down to specific genes of interest. And, rather than focusing on just what is known to be important today, the trends suggests other potential foci of interest. Finally, the connections highlight potential interactions of interest. These approaches can be applied more broadly, for example, using the following additional searches:

What microRNAs are associated with melanoma?
What lncRNAs are associated with melanoma?

Time matters! As you will see from examples like this, you can obtain much deeper discovery and more important insights in a very short amount of time with Qinsight.

1.2 Comparison to PubMed

- The query **What genes are associated with melanoma?** is not understood by PubMed with this "plain English" query giving only about 3000 results.
- With the query "genes melanoma", now you get about 9000 results (less than 15% of the relevant documents), and there is no reasonable way to actually answer the question of "what genes are associated with melanoma" and get to the underlying facts.
- It is hard to see why many of the 9000 results are relevant.
- To get the same answer as Qinsight, the user would have to execute tens of thousands of queries (separate queries for each gene, and they would have to know all the gene aliases. Qinsight does this automatically).

1.3 Comparison to Google Scholar

- The example query gives about 1.2 million results.
- There is no easy way to explore the results. In fact, more results actually translate to more frustration and longer effort for you to find what you really want to know.

More results are not better. More relevant results are what you need.

1.4 Comparison to ScienceDirect

- What genes are associated with melanoma? gives about 59,000 results.
- Impossible to get an answer to "what genes".

2 Example: TP53 and oropharyngeal cancer

With the expanding importance of personalized (or precision) medicine, being able to effectively search for genetic information is critical in all aspects of literature discovery. It is not just early stage target discovery for which genes are important!

In this section, the example search is for **TP53 and oropharyngeal cancer**. TP53 is the official gene name for a tumor suppressor gene involved in more than half of human cancers. Let's get an understanding of its connection to oropharyngeal cancers.

2.1 Qinsight

With Qinsight, this search query returns almost 900 Focused Results. These are the most relevant documents where the author has tied TP53 directly to oral and pharyngeal cancers. [The over 5,000 Broader Results expand the list to include documents that mention both TP53 and any oropharyngeal cancer, but without the explicit or implicit direct connection between the terms.]

From the Focused Results, you can see not only that there is a connection but also the context – for example, p53 is overexpressed and p53 is a prognostic indicator. And, by looking at the Key Concepts, you can quickly see that human papillomaviruses are involved (as well as the different types of oropharyngeal cancer where TP53 is important).

Note that the query used the official gene name TP53, but Qinsight found all the numerous variants of the name including p53 (the protein from the TP53 gene), TRP53, etc. This highlights how Qinsight ensures deeper, more complete and meaningful discovery of genetic information.

> **Focus on what you need.** Depending on your needs, you may find it useful to focus on articles about TP53 and oropharyngeal cancer in people. To do that, add the term **$Clinical** (a *Power Term*™ for finding everything clinically-related) to the Also Containing filter. Now you can immediately focus on this subset. Or, try **$Preclinical** to focus on work in cell lines and animal models.

2.2 Comparison to PubMed

With PubMed, the query **TP53 and oropharyngeal cancer** returns only 45 results – only about 5% of what Qinsight found. A big part of the problem with PubMed is that a search for TP53 only finds TP53, missing the large number of other gene names used in the literature as well as the protein name p53. Let's compare some of these others separately

- **TP53 and oropharyngeal cancer** → 45 results
- **p53 and oropharyngeal cancer** → 227 results
- **TRP53 and oropharyngeal cancer** → 67 results
- **BCC-7 and oropharyngeal cancer** → PubMed reports that BCC-7 was not found in PubMed. In fact, BCC-7 was the original name for this gene, which was the seventh protein identified in basal cell carcinoma cells

This points out that you cannot assume a site like PubMed is giving you everything need when you use a gene or protein name you know. Even the best result is only about 25% of the available literature! To get a better picture, you would have to build a complicated Boolean query, such as TP53 OR p53 OR TRP53 OR …, assuming you even know all the names for this gene that appear in the literature (and there are a lot!).

By the way, since Qinsight's content is much broader than PubMed, so you may wonder to what extent that explains the PubMed results. For the Qinsight results, by clicking the Journal Article Publication Type filter, you can see there are 842 results from journal articles. So, the lower number of results on PubMed truly reflects poor discovery. Plus, keep in mind that PubMed finds articles that mention TP53 and oropharyngeal cancer without regard to whether these terms are connected in a meaningful way.

2.3 Comparison to Google Scholar

Google Scholar also does not "understand" genetics. And, the results obtained in very large part do not directly tie TP53 to the cancer; rather the articles happen to mention both terms.

- **TP53 and oropharyngeal cancer** → 18,900 results
- **p53 and oropharyngeal cancer** → 22,500 results
- **TRP53 and oropharyngeal cancer** → 327 results
- **BCC-7 and oropharyngeal cancer** → 21 irrelevant results

2.4 Comparison to ScienceDirect

ScienceDirect, too, does not "understand" genetics. And, the results obtained in very large part do not directly tie TP53 to the cancer; rather the articles happen to mention both terms.

- **TP53 and oropharyngeal cancer** → 408 results
- **p53 and oropharyngeal cancer** → 1,800 results
- **TRP53 and oropharyngeal cancer** → 12 results
- **BCC-7 and oropharyngeal cancer** → 154 irrelevant results

3 Example: PANDAS

In the biomedical literature, many terms that look alike have very different meanings. In this section, let's explore one example of this problem.

PANDAS in the biomedical literature is Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcal infections. This is literally a different beast from **pandas** – those adorable Chinese bears known also as great pandas (Ailuropoda melanoleuca) or red pandas (Ailurus fulgens).

Abbreviations: There are many different meanings to most abbreviations. For example, MS is used for multiple sclerosis, mass spectroscopy, and many more. So, in general, you should spell out what you are looking for. Qinsight knows within a paper what the abbreviation stands for and will match your query term to the appropriate abbreviation. But, when you use only the abbreviation, it will find all the possibilities – great for when you want to know what the abbreviation might mean.

3.1 Qinsight

In Qinsight, a search for **PANDAS** gives about 400 papers, and we can easily explore the literature about this disease.

A search for **pandas** finds the cuddly creatures. In short, Qinsight – like Yogi – is "smarter than the average bear" (or search engine).

3.2 Comparison to PubMed

- **PANDAS** → 568 results without distinction to the difference between PANDAS and pandas.

3.3 Comparison to Google Scholar

- **PANDAS** → over 34,000 results without distinction to the difference between PANDAS and pandas.

3.4 Comparison to ScienceDirect

- **PANDAS** → over 17,000 results without distinction to the difference between PANDAS and pandas.

4 EXAMPLE: NO asthma

Nitric oxide (NO) is an important cell signaling molecule involved in many diseases. 20% of the literature about nitric oxide uses only "NO" and never writes out "nitric oxide". So, if your search engine cannot search for, or recognize when, NO means nitric oxide, you are guaranteed to miss critical information.

4.1 QINSIGHT

Qinsight is aware of biology and chemistry. As such, it recognizes NO means nitric oxide. And, you can see from the Key Concepts that Qinsight also connects nitric oxide to related concepts including nitric oxide-related genes.

4.2 COMPARISON TO ALL OTHER SEARCH ENGINES

- None of the other search engines understands NO.
- PubMed and ScienceDirect simply ignore "NO" in queries (although PubMed did put in a special case if you query just for "NO" with no other terms).
- Google Scholar finds NO, but with no understanding of what it means. For example, Google finds No. (number), NO (author initials and last names), and no (the negative).

5 EXAMPLE: sapanisertib

Like genes, chemicals – including drugs – have different names and your search engine should work with whatever name you are familiar with. The problems with chemical name recognition are exasperated with early stage entities, such as **sapanisertib**.

5.1 QINSIGHT

In Qinsight, a query for **sapanisertib** recognizes this drug name AND its various code names, providing over 211 results. Users can also immediately glean from the Key Concepts that sapanisertib is an inhibitor or MTOR and then follow up on more details.

5.2 COMPARISON TO PUBMED

- Only 3 results, so the vast majority of the literature is missed

5.3 COMPARISON TO GOOGLE SCHOLAR

- Only about 50 results, so the majority of the literature is missed

5.4 COMPARISON TO SCIENCEDIRECT

- Only 7 results, so the vast majority of the literature is missed

6 EXAMPLE: $Chemicals prevent kidney stones

In this section, let's investigate the use of verbs in your queries. With traditional search engines we have been "trained" to avoid verbs in a query since these terms just add noise to typical results. With the example query **$Chemicals prevent kidney stones**, that noise is a problem since "prevent" and "kidney stones" are not necessarily tied together in a meaning way.

6.1 QINSIGHT

Qinsight is powerful enough to understand actions and to find meaningful connections within the literature.

In this example, we are interested in knowing what compounds might be used to prevent diabetes or its complications. $Chemicals is a *Power Term* shortcut, which instructs the application to search for actual compounds, not the generic term chemical (and related general concepts). [Or, you could use plain English, as in What chemicals prevent diabetes?]

In the Qinsight results, you can immediately see how certain chemicals are working to prevent kidney stones (nephrolithiasis). The Key Concepts shows all these compounds for easy exploration.

6.2 COMPARISON TO PUBMED

- Using **chemicals prevent kidney stones**, there are no results. ☹ If you want to investigate this further, you will have to do a whole lot of searches, each for a different compound – an impossible task.

6.3 COMPARISON TO GOOGLE SCHOLAR

- Using "**chemicals prevent kidney stones**" (since Google does not support *Power Terms*), there are about 35,000 results, many of which do not connect "prevent" with "kidney stones". Good luck getting a comprehensive answer to the question.

6.4 COMPARISON TO SCIENCEDIRECT

- Using "**chemicals prevent kidney stones**" (since Google does not support *Power Terms*), there are about 13,500 results, many of which do not connect "prevent" with "kidney stones". Here, too, it will be very difficult to answer the question posed.

7 EXAMPLE: lateral sulcus

Like genes, chemicals, and diseases, parts of the anatomy are also known by many names in the literature. The lateral sulcus is no exception. No matter which name you use (**lateral sulcus**, **Sylvian fissure**, **sulcus lateralis cerebri**, etc.), you should expect to find all the relevant literature.

7.1 Qinsight

With Qinsight, the query lateral sulcus gives about 3,000 results. As you would expect, any other name for this anatomical structure gives the same results.

7.2 Comparison to PubMed

- **lateral sulcus** → about 2,000 results
- **Sylvian fissure** → about 1,200 results
- **sulcus lateralis cerebri** → 0 results

7.3 Comparison to Google Scholar

- **lateral sulcus** → over 130,000 results (many of which include "lateral" and "sulcus", but not "lateral sulcus")
- **Sylvian fissure** → about 38,000 results
- **sulcus lateralis cerebri** → about 3,500 results

7.4 Comparison to ScienceDirect

- **lateral sulcus** → about 41,000 results
- **Sylvian fissure** → about 8,000 results
- **sulcus lateralis cerebri** → about 200 results

8 Example: What methods detect metastases?

Techniques and methods are an important part of understanding the literature and finding what you need. In this section, we use the query **What methods detect metastases?** to demonstrate differences between search engines.

8.1 Qinsight

For this query, Qinsight finds almost 3,000 focused results that directly tie different methods to metastasis detection. And, the different methods are listed in the Key Concepts filter. This focused set of results is complemented by over 230,000 broader results.

Note that with over 10.5 million of the 40+ millions authoritative documents in Qinsight being full-text, you are covering many Materials and Methods sections. And, Qinsight is rapidly expanding the number of full-text articles that are not available via Open Access.

8.2 Comparison to PubMed

- Only about 5,000 results
- No way to easily identify the actual methods

8.3 Comparison to Google Scholar

- Over 900,000 results
- Results include the word "method" instead of actual methods, so papers that have a header call "Methods" section are included even if they are not about metastasis detection methods.

8.4 Comparison to ScienceDirect

- Over 85,000 results.
- No way to easily identify the actual methods

9 Example: children

Effective discovery from the literature requires a conceptual approach. Simply looking for a certain keyword is almost certainly going to miss critical information. This problem encompasses the ability of the search engine to recognize stemmed forms of terms, but it is a much broader problem that just stemming (or lemmatization).

In this section, we will compare search engines in finding the concept of **children**.

9.1 Qinsight

Qinsight understands concepts, not just words. This is critical for finding all the relevant information.

For example, if you are searching for **children** (alone or as part of a more detailed query), you should find the same results using **child** or **children** or even **childhood**. If fact, with Qinsight's powerful concept basis, you will also find "juvenile", "kids", "boys", and "girls" – all of which are also conceptually children. Boys and girls are, or course, subsets of the broader concept of children.

9.2 Comparison to PubMed

- **children** → about 2.25 million results
- **child** → about 2 million results
- **childhood** → about 200,000 results

and related concepts such as juvenile, infant, and boy are missed. It is nearly impossible to assess the degree to which these result sets overlap.

> **Use of MeSH Terms:** With PubMed, you can use the appropriate MeSH terms (or simpler, the Ages filter) to limit results to articles that have specific age ranges. But, this approach does not assure that any specific age range is, in fact, connected with the rest of your search terms. Also note that many MeSH terms are not added until well after the article first appears in PubMed, so searching by MeSH may not cover recent articles.

9.3 Comparison to Google Scholar

- **children** → about 4.3 million results
- **child** → about 3.3 million results
- **childhood** → about 2.9 million results

and related concepts such as juvenile and boy are missed. It is nearly impossible to assess the degree to which these result sets overlap.

9.4 Comparison to ScienceDirect

- **children** → about 1.3 million results
- **child** → about 850,000 results
- **childhood** → about 400,000 results

and related concepts such as juvenile and boy are missed. It is nearly impossible to assess the degree to which these result sets overlap.

10 Visualizations

Qinsight is unique among all tools in having **visual analytics based on the text content** (as opposed to simple bibliometrics). These visualizations provide insights that are simply not possible from looking at lists of results.

Concept Cloud, which contains concepts specifically related to the query, provides a visual summary of what the results are about. This information cannot be gleaned from a simple list of results.

Concept Trends is another way to see all the concepts related to the user's query, but also predicts trends. These trends cannot be discerned from a list of results.

Concept Connections provides a means to identify meaningful connections among Power Term members. Such connections are very difficult to find in simple lists of results.

Category Map enables you to explore the results by categories, such as Industry, Practice Area, and more.

11 Conclusions

Decisions based on information gleaned from literature discovery impact every aspect of the healthcare industry. Whether it is multimillion-dollar decisions made by pharmaceutical companies, efforts to gain funding in the academic environment, or treatment choices for patients, you deserve smarter, faster literature discovery solution. Qinsight can make you a hero in your organization.

2505 Anthem Village Dr. | Suite E170 | Henderson, NV 89052

Toll-free: +1-844-44 BIO AI (+1-844-442-4624)
Email: info@quertle.com