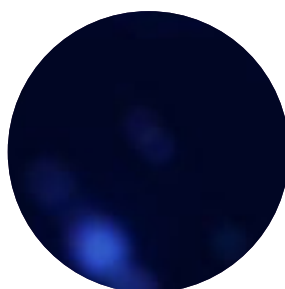● **Whitepaper**

Transforming Research Results:

The Power of the MDS Approach

Introduction

Evalueserve is a leading player in the IP and R&D search and intelligence business, with more than 22 years of experience. Evalueserve's IP and R&D team of over 300 analysts is one of the largest in the world. Evalueserve has conducted millions of hours of patent and technical literature searches, and this report is based on that experience. This white paper shares insights on the importance of data collection in research and how adopting a multiple database strategy improves the accuracy and reliability of research results and the quality of the study.

The Importance of Data Collection in Research

At Evalueserve, we understand that data collection is the crucial aspect of research that ensures accurate and precise results. The quality of data collection has a direct impact on the reliability and scope of the study, highlighting the importance of data collection in determining the overall quality of the study. We employ various data collection methods, including searches on curated commercial databases (patent and non-patent), established search engines (Google, Bing, Yahoo, and others), AI-based tools (domain-specific and foundational), surveys, interviews, and observations to collect comprehensive, accurate, reliable, and relevant data. We ensure that our data collection process is meticulous, detail-oriented and methodical to ensure good practices.

Our Research Methodology

In the 22 years we have been in the intellectual property search and analysis market, we have found significant discrepancies and variations in all databases. To ensure accuracy and reliability of research results, we have adopted a process of using more than one database to search and retrieve the same type of data type for a study. Searching multiple databases allows us to compare data on the same topic from different sources, identify and correct discrepancies or errors, and triangulate data to validate or challenge findings, ultimately increasing the accuracy and reliability of research findings.

Understanding the Recall with Multiple Databases

Using multiple databases in patent searches is an effective strategy that can greatly improve the scope and accuracy of search results.
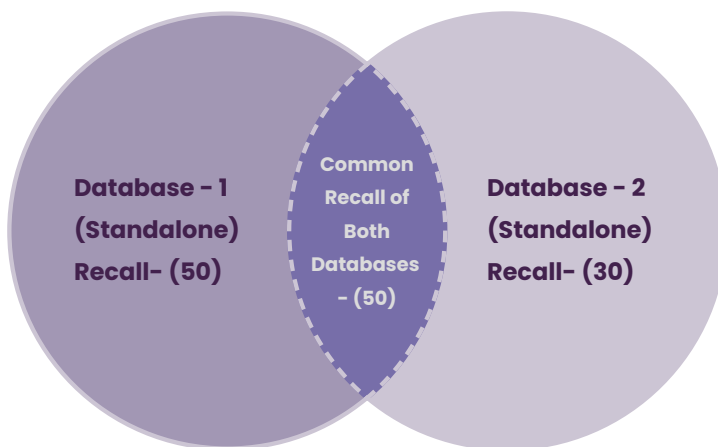
Consider two patent databases, 1 and 2, that contain patents on a particular technology. Using a Venn diagram, we can visualize the number of patents in each database and the number of overlapping patents in both databases (see Figure 1). By using two databases, we can retrieve the entire set of patents from both databases, making the search result more comprehensive.

However, there may be more efficient uses for a multiple database strategy. We can focus on the unique patents in each database to increase the discoverability of the search result.
By selecting the data that exists in the intersection of the two databases, we can achieve a more precise search result that is relevant to the search topic and current in both databases. This approach can also reduce the risk of retrieving irrelevant patents that can take time to analyze.

In addition, adding more databases to this approach can further increase the search result's coverage, recall, and accuracy of the search result, allowing researchers to select the dataset that best fits the needs of their research study.

Database – 1 (Standalone) Recall- (50)

Common Recall of Both Databases – (50)

Database – 2 (Standalone) Recall- (30)

Fig. 1: Coverage from 2 Databases for Patent Search

Using Multiple Databases for Research

At Evalueserve, we have conducted several studies to test the completeness of data for patent and non-patent searches. The study we use in this paper was conducted over a three-year period with the goal of comparing the performance of two different patent databases in terms of their ability to perform effective patent searches and retrieve a comprehensive patent dataset. To achieve this, a unified search strategy was used, first established in one database and then replicated in the other patent databases.

In order to implement the search strategy in multiple databases, we had to convert the search strategy into a format that was compatible with the syntax and operators of the second database. This required modification of the search strategy to account for differences between databases. We also used the additional curated data fields from the second database where appropriate. However, the core elements of the search strategy, such as keywords, synonyms, and classifications, remained consistent. In this way, we were able to evaluate the databases on a level playing field without search strategy bias. Overall, our study aimed to provide a comprehensive evaluation of various patent databases in terms of their ability to perform effective patent searches using a consistent search strategy.

Biases with using Multi Databases Strategy

While we have made efforts to conduct the comparisons between databases on a level playing field, we are aware that there will always be differences in the results of the various databases due to the inherent biases of these databases and the biases in the procedures developed. We have listed some of the key considerations as following:

• **Differences in database structure:** Different databases may have different fields and categories for storing patent information, which may affect the accuracy and completeness of search results.

• **Differences in indexing:** Databases use different algorithms to index and retrieve patent information, which may affect how search terms are interpreted and matched with relevant patents.

• **Linguistic differences:** Different databases may use different natural language processing techniques, which may affect the way search terms are understood and matched with relevant patents.

• **Synonymy and polysemy**: Different databases may have different approaches to handling synonyms and polysemy terms, which may affect the accuracy of search results.

• **Differences in patent coverage:** Databases may have different levels of coverage for different types of patents or patent applications, which may affect the completeness of search results.

• **Time differences:** Databases may have different time periods for which they have data, which may affect the completeness and accuracy of search results. Additionally, not all databases are updated at the same time.

• **Curated Data Fields:** Databases provide additional curated data fields for search, which can affect the number of search results obtained from search. There are two schools of thoughts around this, one believes in data purity even at the cost of comprehensiveness while the other focuses on comprehensiveness at the cost of data purity.

- **Custom-made search strategy for the first database:** The original search strategy may have been tailored to the specific search algorithms and indexing methods of the first database and may not be appropriate or effective in the second database.

- **Modification of operators and syntax for the search strategy used in the second database:** The substitution of operators and syntax may not accurately reflect the intended meaning or context of the original search string, leading to inaccurate or incomplete search results.

The Multi Database Strategy Experiment and its Findings

Over the past three years, we have conducted an experiment using the same search strategy in two leading commercial patent databases. Our global research teams have applied this approach to over 2,000 studies to compare the search results of the two databases and determine the strengths and weaknesses of the two databases for different types of studies, including intelligence studies such as landscapes, technology scouting, competitive intelligence, alerts, FTOs, infringement, and more.

Distribution by Study type
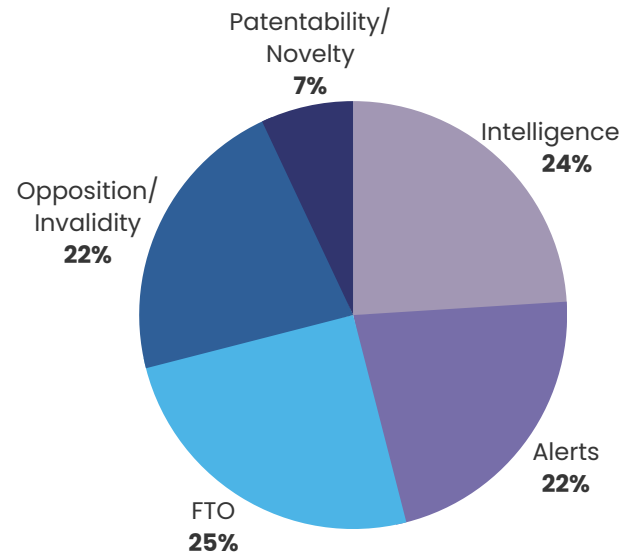
Figure 3: Percentage distribution of study types considered for this experiment.

The research team systematically conducted the search strategy in both databases and analyzed the results to determine which database provided more relevant search results for various studies. The team determined the strengths and weaknesses of each database in terms of coverage, recall, precision, and quality (see Figure 4).

The results of the experiment showed that each patent database has its own advantages and disadvantages, and that the choice of database for a particular study depends on the specific needs of the study. However, the combined use of multiple databases offered significant advantages over the use of a single database. These advantages included retrieval of a more comprehensive set of patents and improved retrievability.

Comparison of 2 Databases
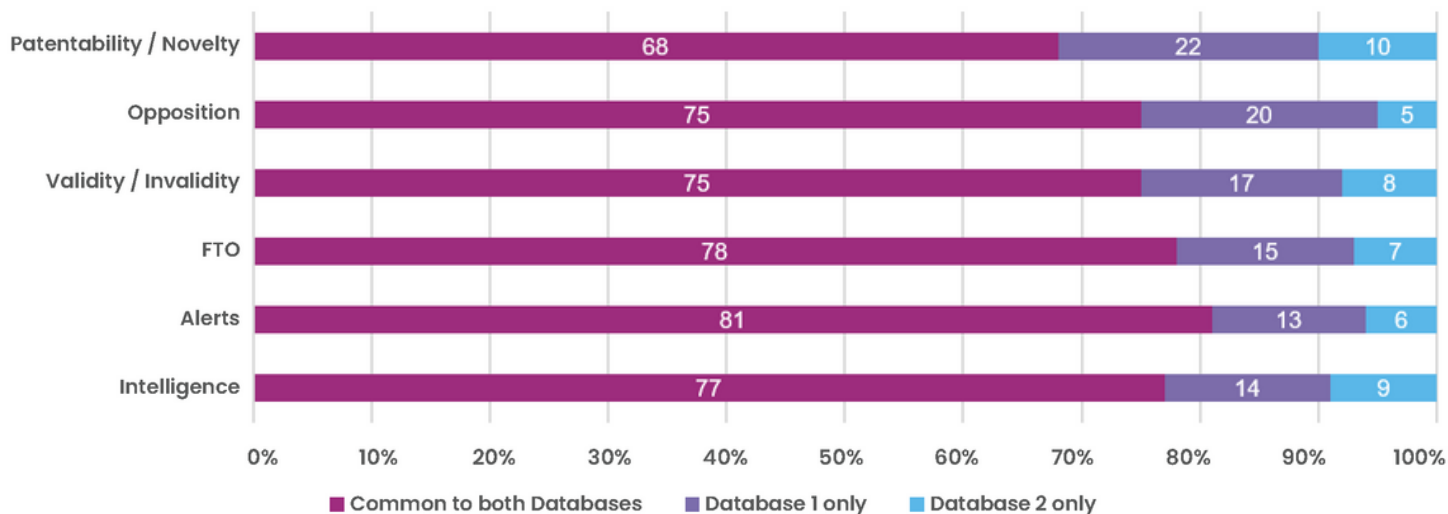
Figure 4: Percentage of results uniquely obtained from each database and the percentage of results common to both databases for different types of studies.

The findings of the experiment have enabled us to provide clients with more comprehensive and focused search results, thereby improving the overall quality of the service. The experiment also underscored the importance of having a defined and structured set-up for conducting searches using a multi-database strategy. Such a set-up would require analysts who possess a high level of skill in the art of searching and possess strong data handling capabilities. To ensure compliance with the multi-database strategy, it is important to periodically measure the outcomes and adjust on multiple fronts.

This includes determining the key databases to consider for the search, framing an appropriate search strategy, and selecting the best approach for searching based on the specific requirements of the study.

Overall, the experiment demonstrated the importance of having a structured and well-defined process for conducting patent searches using a multi-database strategy. By implementing such a process, organizations can deliver more comprehensive and targeted search results to their customers, ultimately improving the quality of their services.

Evalueserve MDS Team Set-up for Patent Data Collation for Studies

Evalueserve IP and R&D is committed to improving the quality of its studies, combining data compliance with efficiency and cost optimization. Driven by the findings of the above experiment along with many others, to achieve this, we have established a dedicated Multi Database Strategy (MDS) team specialized in conducting patent searches in multiple databases, inspired by the results of the above experiment and many others. We currently use five (5) commercial databases in our internally deployed MDS process.

The MDS team consists of experienced IP searchers who have extensive experience in conducting searches in various patent and non-patent databases.

The figure below shows process chain of MDS set-up.

They are very familiar with the syntaxes, search fields and functions of the various databases. Their main task is to execute the search strategies developed by individual searchers or search teams for one database in several databases and to compile the search results in a systematic and standardized way.

The MDS team collaborates with other groups, such as the training, automation, digital transformation, and engineering teams, to provide a more comprehensive service to customers. The establishment of the MDS team has resulted in numerous benefits for customers, searchers, and led to cost optimization, resulting in a more efficient and comprehensive business model.
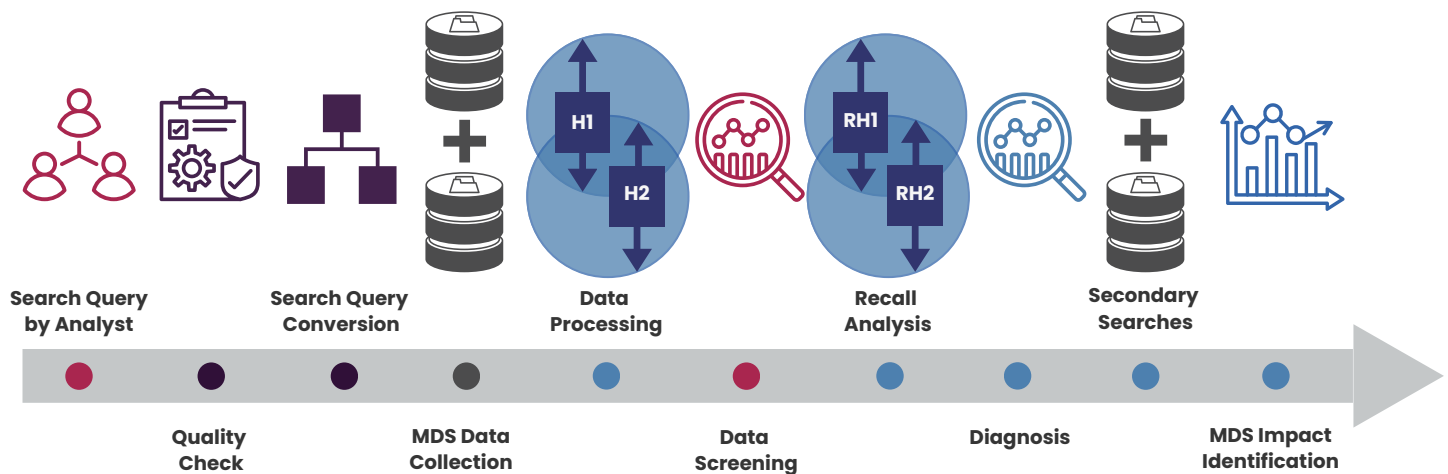

Fig. 2: The MDS set-up process chain

Conclusion

Our study highlights the importance of data collection in IP and R&D research and underscores the need for a comprehensive and thorough process. Each patent database has its own strengths, weaknesses, and biases which can impact the way search results are retrieved and shared with users. To ensure the best interests of clients and obtain accurate search results, it is essential to conduct searches through multiple databases.

However, dealing with the vast amount of data generated by multiple database searches can be overwhelming. To overcome this challenge, it is crucial to have well-defined processes and a highly skilled team to manage the searches.

Evalueserve IPR &D business unit is a pioneer in this regard and was the first in the sector to establish a dedicated Multi Database Strategy Team with access to multiple commercial databases. While building such a team and having access to commercial databases is a cost-intensive process, the benefits of this approach are significant. These include increased quality in our studies, comprehensive results, and benchmarking of databases and data sources.

Learning and Recommendations

Based on the above, the five key recommendations and learning points for using a multi-database strategy to conduct patent searches are:

• **Improved search results:** Conducting patent searches in multiple databases can lead to a more comprehensive set of search results covering a broader range of relevant patents. This can improve the accuracy and efficiency of the search process and lead to better insights and analysis.

• **Overcoming inherent biases:** Each patent database has its own strengths, weaknesses, and biases that can affect search results. Conducting searches across multiple databases can help overcome these biases and provide a balanced and unbiased view of the patent landscape.

• **Defined processes and specialized teams:** To effectively manage the vast amounts of data generated by searches across multiple databases, it is important to have well-defined processes and a specialized team with highly skilled analysts who are well-versed in the nuances of various databases.

• **Benchmarking of databases:** Using a multi-database strategy can help benchmark different databases against each other and identify the strengths and weaknesses of each. From this, future search strategies can be derived and the search process can be optimized.

• **Cost optimization:** While building an MDS team and accessing databases can be a cost-intensive process, the benefits of this approach can ultimately lead to cost optimization by producing more accurate and comprehensive results, requiring fewer additional searches, and reducing the risk of missing relevant patents.

If you would like to discuss this report or our approach, or for more information, please contact **iprdsolutions@evalueserve.com**

About Evalueserve

Evalueserve combines insights emerging from data and research with the efficiency of digital tools and platforms to design impactful solutions. A global team of 5,000+ experts collaborate with clients across 15+ industries.

Our proprietary technology platform **Insightloupe** is a powerful analytics engine that combines the best in human expertise with cutting-edge technology to offer unparalleled patent and non-patent landscaping insights. With Insightloupe, you can access expert-backed IP and R&D insights, allowing decision-makers to make informed and accurate choices with precision and clarity.

Find out more

Searchstream is a breakthrough Intellectual Property Search ordering platform, amplified with domain-specific AI. Searchstream's efficient patent search platform allows clients to select the scope of search, the number of databases used (promoting a stronger search recall), and the specific data types they would like to have included in your report. In other words, Searchstream puts you in the driver's seat.

Find out more

Evalueserve Disclaimer

The information contained in this report has been obtained from reliable sources. The output is in accordance with the information available on such sources and has been carried out to the best of our knowledge with utmost care and precision. While Evalueserve has no reason to believe that there is any inaccuracy or defect in such information, Evalueserve disclaims all warranties, expressed or implied, including warranties of accuracy, completeness, correctness, adequacy, merchantability and / or fitness of the information.
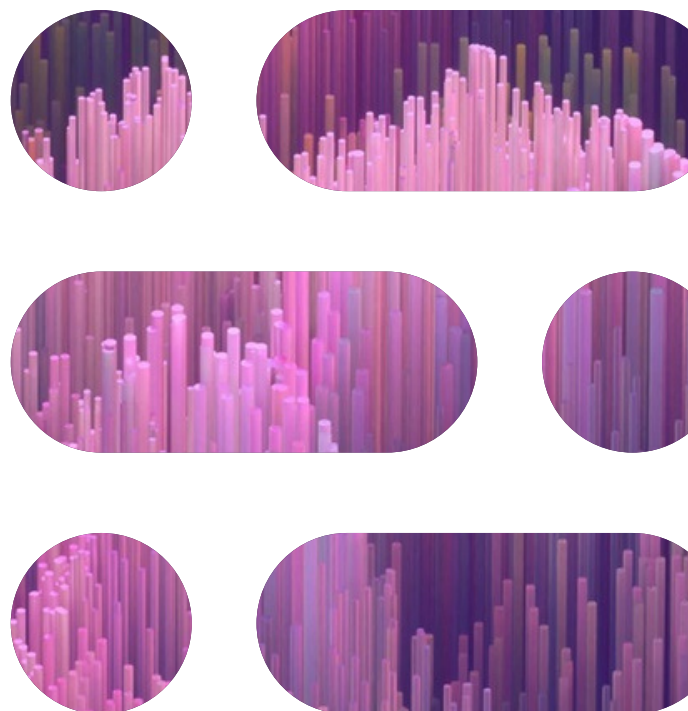
evalueserve. com

Evalueserve's Contacts
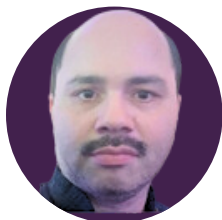
Dilip Kumar Upadhyay

Associate Director, Digital Transformation Lead, IP and R&D, Evalueserve
Dilip has over 15 years of experience of leading and managing various teams across IP and R&D LOB including Med-Tech and Healthcare team, MDS team, Incubator team. Currently, he leads the Digital Transformation for IP and R&D, including adoption of Generative-AI across workstreams.

Yogesh Shinde

Senior Consultant, Lead for MDS team, IP and R&D, Evalueserve
Yogesh has more than a decade of experience managing complex IP searches and patent alerts for different fortune 500 clients in Chemical and Chemistry space. Over the last 2 years, he is leading the MDS team, and ensuring that IP and R&D teams always have best available data available for their projects and studies.