



Children's Memorial Research Center: Speeding the Search For Brain Tumor Treatments with Grid-Enabled Text Mining

United Devices Case Study
All Industries: Text Mining & Data Analytics

Solution Overview

Customer Challenge

Children's Memorial Research Center (CMRC) is one of only five research institutions in the country devoted exclusively to pediatrics. Its groundbreaking programs and leading investigators attract millions in federal research dollars. In the search to identify gene therapies for pediatric brain tumors, CMRC's Dr. Eric Bremer needed an efficient way to scan hundreds of thousands of journal articles to uncover key gene relationships, gain new insight into underlying tumor pathology, and ultimately predict which genes might lead to successful treatments.

Solution

After utilizing SPSS' enterprise data mining workbench Clementine® to analyze and classify tumor data, Dr. Bremer turned to SPSS' linguistics-based text mining application LexiQuest Mine™ to explore the sea of published research for specific concepts and relate the analysis to known diagnosis criteria. He then implemented a grid solution to accelerate processing without expanding the Center's hardware infrastructure. Grid-enabled LexiQuest Mine has made it possible for Dr. Bremer's team to effectively work with over five years of published research to build a knowledge base that will help improve the speed and accuracy of brain tumor treatment development.

Benefits

CMRC found that by combining a robust text-mining application with a grid solution for processing, they achieved:

- **94% timeframe reduction** from 24+ hours to just over an hour
- **Reduced cost of analysis** by avoiding hardware investments
- **Expanded research scope** with the ability to run more jobs a day

Dr. Bremer's team has successfully defined workflows and automated nearly 80% of this process with the grid in place.

Technology

- Grid MP™ platform from United Devices
- LexiQuest Mine™ from SPSS

Children's Memorial Research Center: Applying Innovative Approaches to Research & Diagnostics

Established in 1986 as a formalized basic science research program, Children's Memorial Research Center is the research arm of Children's Memorial Hospital, the pediatric teaching hospital for Northwestern University's Feinberg School of Medicine and one of only five institutions in the country devoted solely to pediatric medicine.

CMRC's vision is to become the preeminent child health research organization dedicated to improving the health of children nationwide. The Center's ability to attract leading investigators has considerably advanced its goal of ensuring scientific knowledge is translated into tangible and effective clinical uses for the benefit of all children.

CMRC has organized its work around seven interdisciplinary research programs, including Cancer Biology and Epigenomics. As Director of Brain Tumor Research within this program, Dr. Eric Bremer is charged with finding new, innovative ways to identify genes that might lead to successful treatments.

Know Your Foe: The Pediatric Brain Tumor

Pediatric brain tumors, the most common type of solid tumor in children, are also the most fatal child cancer. Half of those diagnosed -- nearly 3,000 a year in the U.S. alone -- will die within five years. And for survivors the prospects are sobering: neurological disabilities, retardation, and psychological problems are among the long-term problems these children can suffer.

Once a child is diagnosed with one of the 12 pediatric brain tumor types, the options are few. Outside of surgery (which is risky) the treatments available are rare. That's

"There are plenty of scientists out there publishing results that could be critical for our work. But because these results are often buried in studies that don't immediately seem relevant, key pieces of data remain untapped. We needed a way to take advantage of all this published work, to sift through and uncover only the results that are relevant to our work -- no more, no less."

Dr. Eric Bremer
Director of Brain Tumor Research,
Children's Memorial Research
Center

why Dr. Bremer and his team at CMRC are leading the search for new therapeutic targets -- key genes associated with specific tumor types that should be examined for therapeutic potential.

"If we can better understand which genes are related not only to the disease itself but also to fundamental biological processes (like growth and development), we can better target certain genes for further research,"

explains Dr. Bremer. "If a gene or cluster of genes has been proven to affect certain disease pathways, for example, it might be possible to develop that into a treatment for that particular pathway."

Dr. Bremer and his CMRC team had already built a database of tumor-related gene information using the SPSS data mining workbench Clementine. Now, they needed to evaluate that information against what was known about how genes interact in other, similar situations in order to uncover relationships that might help in their search for tumor therapy drug candidates.

A Global Collaboration

Fortunately, Dr. Bremer and his team were not starting from scratch. Thousands of researchers worldwide have published studies in professional journals that reference genes and their relationships to diseases, treatments, and each other. Publications like the *Journal of Medicinal Chemistry*, *Journal of Biological Chemistry*, and *Science* are valuable tools enabling researchers worldwide to share results that may further each others' studies.

Dr. Bremer recognized this body of work as a potential source for valuable insights into new therapeutic approaches for pediatric brain tumors. By uncovering insights hidden within the pages and pages of research, Bremer and CMRC can better predict which genes to target for development.

But first you have to know where to look.

Needles in Haystacks

Unfortunately, there are a limited number of free hours in a scientist's day, and when picking articles to review in search of valuable insights an article is likely to be overlooked if the primary topic doesn't seem relevant.

As Dr. Bremer explains, "I wouldn't normally scan an article on fruit flies or bacteria, for example, to learn about genes associated with pediatric brain tumors. But by ignoring these studies we may be missing out on any number of insights about potentially therapeutic genes."

Dr. Bremer's team needed a way to leverage over 125,000 specific articles from 21 journals and turn this data into a useful tool for research. But the challenge of wading through literally millions of pages of text required a manageable and cost-effective solution.

"LexiQuest Mine has been a state-of-the-art application for gigabytes processing. With United Devices, the era of text mining terabytes has begun."

Olivier Jouve
Vice President of Product
Marketing, Data and Text Mining
SPSS

Traditional keyword search solutions tend to produce too few – or too many – hits and do not help scientists zero in on the few results that can help provide critical information about a particular gene association and its biological relevance. So, without tools to identify and extract the right published data, much of the information that can help Dr. Bremer and scientists like him remains hidden in the overwhelming volume of published research.

LexiQuest Mine: Extracting Insights, Building Knowledge

To solve the problem of scanning hundreds of thousands of journal articles for relevant information, Dr. Bremer and his CMRC team turned again to SPSS and their text mining application LexiQuest Mine.

LexiQuest Mine works by employing a combination of dictionary-based linguistic analytics and statistical proximity matching to identify key concepts, including multi-word concepts. Based on a linguistic analysis of the context and semantic nature of the words, LexiQuest Mine is able to identify the type (organization, product, genes, etc.) as well as the degree of the relationship between the words and other concepts.

LexiQuest Mine has two operating modes that can be separately or subsequently used:

- Extraction of concepts, based on Part of Speech (PoS) tagging
- Text Link Analysis, to identify semantic relationships between known concepts

Ultimately, the results can be displayed in a color-coded graphical map so that the analysts can clearly identify relationships.

"This innovative implementation shows how broad the application base for grid within enterprises is becoming. Our customers are finding new uses for their grids every day – from managing workloads across disparate clusters to accelerating specific software tools. With Grid MP, organizations are able to innovate faster and make better decisions – for strong competitive advantage."

Jikku Venkat
CTO
United Devices

Using LexiQuest Mine, Dr. Bremer was able to detect protein or gene interactions that were identified across a large number of scientific publications and relate them to known brain tumor diagnosis criteria catalogued in the Clementine database.

With this tool, Dr. Bremer's team would now have the ability to scan the database of available journals and create a set of data that could advance their research immeasurably.

Challenge: Speeding the Search Process

While LexiQuest Mine made it possible to automate the process of intelligent data searching, the turnaround time for job processing was still prohibitively long. "I quickly realized it wasn't realistic, even with this application, to mine the amount of data I needed to review," said Bremer.

If the text mining process was to be used as a valuable tool for research, then a solution for faster processing would be required.

"The grid solution has literally changed our definition of what is possible. Where before we were able to scan only a few thousand articles a day, now we can examine 100,000 articles in the same amount of time. We can also pull out more concepts and can rerun the data in different ways, rather than just once, to improve the precision and quality of what we learn. It's a dramatic increase in scope that analysts and researchers in other fields are going to want to achieve as well."

Dr. Eric Bremer
Director of Brain Tumor Research,
Children's Memorial Research
Center

Grid MP for Optimized Processing

To help surmount the obstacle of extended processing times, SPSS introduced Bremer to United Devices (UD), a leading provider of enterprise grid computing software. SPSS wanted to maximize the benefits of its LexiQuest Mine solution for CMRC and offered to grid-enable LexiQuest Mine on a United Devices Grid MP platform.

"The Grid MP solution was attractive as an alternative to investing in high-end servers," explains Dr. Bremer. "With grid computing we were able to use in-house machines for processing rather than investing in new hardware or, worse, reducing the scope of our project."

Grid MP consists of a server and agents distributed to designated machines (servers, clusters, or desktops) that can be in the same room or on multiple continents. The software schedules processing jobs to machines that "advertise" their availability (meaning they have idle processor time and are equipped to perform the necessary computations) – ensuring work is scheduled to the best-suited resource. The machines return results to the server, which recompiles the data and delivers it to the user in an easily searchable database format.

Robust provisioning policies allow administrators to define usage policies and eliminate the need for the researcher to assign the jobs to specific machines.

"From the user's standpoint," says Bremer, "nothing has changed except that we get results in a fraction of the time." Bremer's team is now able to run multiple jobs a day and retrieve more results and thus improving both the timeframe and the scope of their work.

Up and Running in a Snap

The CMRC deployment was typical, taking only a few hours to install and configure. Within days LexiQuest Mine was up and running thanks to Grid MP's flexible application framework. Dr. Bremer was able to start using the grid-enabled system within weeks of initially contacting UD, and benchmarking was complete in under a month.

Value and Benefits

Thanks to the SPSS / UD joint solution, Dr. Bremer has established a growing knowledge base as a critical tool for advancing research on pediatric brain tumors.

Project benefits include:

- **Timeframe Improvements:** Project runtimes fell by 94%. Projects that used to take 24-26 hours can now be run in just over an hour, so CMRC can now execute numerous searches during the workday – a vast improvement over the previous 24-hour wait time for results from a single search.
- **Reduced Costs:** By harnessing the power of existing CPUs within their organization, CMRC was able to avoid an investment in expensive high-performance computing hardware and instead increase the performance of resources already purchased.
- **Expanded Scope of Research:** With the grid-enabled text mining solution in place, Dr. Bremer and his team can harness the full body of published results available to them. They are now able to search hundreds of thousands of articles a day and to refine the way in which they analyze their results.

Broad Relevance

The need to quickly search, process and analyze vast volumes of text and other data is not unique to the healthcare field. Dr. Bremer stresses that this solution would be useful in a wide range of areas, "from lawyers mining patent applications to law enforcement agencies searching for clues in databases of evidence."

The rapid increase in available information over the past decade necessitates approaches like the joint UD / SPSS solution to harness and benefit from that information – as fast as it becomes available.

Future Possibilities

CMRC is already investigating grid-enabling other processes related to this program. For example, Dr. Bremer's team has enabled the automated data transfer

tool, GETITRIGHT by CTH

Technologies, used for accessing journal articles and preparing them to be mined. This solution automatically connects and downloads full-text journals from the web and does the necessary processing to deliver output that is readily available for text mining.

Dr. Bremer is also looking for ways to make these resources available to others. His knowledge base will be valuable to any doctors seeking to classify tumors more quickly and accurately, and the applied grid procedure would be relevant to researchers in a variety of fields. So others may benefit from his efforts, Dr. Bremer has published and presented his results in a variety of forums.

Lessons Learned

- **Consider new approaches to manual tasks:** The prospect of manually reviewing thousands of journal articles for clues to tumor therapy was daunting. But by implementing automated tools for text mining, Dr. Bremer's team made it possible to perform work that was previously thought to be out of scope.
- **Promote new technologies across the business:** The success of CMRC's initial SPSS implementation led to the search for (and discovery of) additional cost and time-saving uses for grid. Companies should seek ways to apply technologies for benefit in multiple areas of business.
- **Challenge the boundaries of what you thought was possible:** Examine the limits of what was possible in the past and consider what would be required to move beyond those limits. Then consult with trusted and respected experts about how to meet these requirements.

For More Information

- **United Devices:** sales@ud.com or <https://www.ud.com/contact/gridsolutions.htm>
- **SPSS:** www.spss.com
- **Children's Memorial Research Center:** www.childrensmrc.org/

United Devices and Grid MP are trademarks or service marks of United Devices, Inc. Other company and product names may be the trademarks or service marks of third parties. Copyright © 2005 United Devices Inc. All rights reserved.

RESULTS

Reduced Timelines

"Thanks to the grid, we were able to reduce processing time from over 24 hours to just over an hour. We wouldn't have been able to truly leverage this valuable expanse of information without the United Devices solution."

Expanded Scope

"It's changed the way we think about what can be done – since now we don't have to limit the scope of our searches, we can be confident we've left no stone unturned in our research."

Global Collaboration

"Now we can fully leverage the published research of scientists all over the world – this is a big step in achieving a truly collaborative global research community."