



*Coronado Group, Ltd.*

**Intelligent Search for  
Answering Clinical Questions**  
Coronado Group, Ltd. Innovation Initiatives

Search The Way You Think



## **Introduction**

Coronado Group, Ltd. is building an intelligent search tool (IST) that supports enhanced investigative search and information retrieval for radiologists, clinicians, and medical researchers. The first release of the IST is focused on answering clinical questions in the diagnostic imaging domain.

The new Coronado IST is a sophisticated search engine that is able to search and retrieve documents based on the concepts and ideas they contain. A sophisticated concept space that defines all of the concepts in the document and all of the relationships between all of the documents embodying the same concepts is used to execute searches. Users pose a single query that addresses a complex patient question and the IST search engine finds, ranks, and returns the documents that are conceptually and semantically closest to the search source. Searches are executed using the content of a journal article, manuscript, news story, image interpretation, radiology report, text data from a clinical trial, or any combination of these sources as input into the IST. The search engine supports entry of gigabytes of text as a search source. This allows searcher to enter all of the information about a complex issue or a new discovery in a clinical trial to explore the searchable document database. Instead of distilling complex ideas or problems into keywords and Boolean search strings (and, or, not), users simply copy and paste the text or documents that best represent their problem into the search box. This approach allows a clinician or researcher to pose a single search that “understands” the question and receive a comprehensive answer continuing the most relevant peer-reviewed authoritative citations across a comprehensive searchable database.

Our technology finds and indexes all of the concepts embodied in a document not just its main themes. It doesn't rely on keywords or simple word matching to execute a search. It finds information in existing medical literature that may have been tangential to the original research work but is important to a new investigation or patient care issue. This helps harvest new information in the existing work potentially leading to new discoveries and insight.

The technology's ability to find information that is generally “unfindable” using conventional search techniques allows researchers to find new data in patient reports, medical research and publications. This technology has the potential enhance efforts in evidence-based medicine and new comparative effectiveness research (CER) efforts by facilitating assembly of more accurate evidence. This approach to information retrieval is more efficient and more cost effective because it lets users search across the widest possible universe of patient information while returning highly focused, fine tuned results.



Leading medical research institutions, biomedical engineers, academic radiologists and others involved in the emerging multimodality imaging approaches for diagnosis and treatment of patients have called for new search and information retrieval tools that address the growing complexity of data in the field. Biomedical imaging scientists have an increasing need to access interdisciplinary information to make informed decisions about patient care and best practices. The accelerating pace of creation of new digital medical information and the convergence of existing radiology technologies and modalities with new molecular imaging protocols and technologies, and the increased use of multimodality investigations and treatments call for search techniques that can address this increasingly complex information. Coronado's IST is designed for this type of information search.

### **The Challenge -- End Search Paralysis**

The challenge facing researchers today is how to overcome the limitations of conventional search techniques to find better information while providing searchers with enhanced confidence in the results of their searches. Searchers faced with a complex patient care or research question need to find more actionable information across the widest possible spectrum of patient reports and research.

Conventional search techniques leave the vast corpus of medical information inaccessible. Current search engines compel users to devise the "*perfect query*" to find what they are looking for. User need to know and provide the exact words needed to find relevant material. Search software relies on exact matches of keywords, or exact citations to execute effective searches, or a combination of keyword, citation, and metadata tagging like that used in PubMed. Researchers must execute multiple queries using current vocabularies to eliminate search results that contain the same words but do not cover the topic being searched for.

"Search paralysis" sets in when combinations of words yield no new or usable results. Users are forced to begin anew with a new set of terms. The more novel or complex the search topic, the more difficult it is to get meaningful results or to follow a thread of subject matter in the way that a clinician, or referring physician thinks about and defines his/her expertise or the search goal at hand. The result is that the expansive universe of research, clinical data, and best practice information is inaccessible because there are limited ways to find subtle or nuanced information buried within the documents and reports.

Using conventional search techniques, the quality of the search results is contingent on the searcher's understanding of the keywords and vocabulary associated with the specific specialty or subspecialty. A series of Boolean operators (and, or, equal to, not, not equal to) are needed to build searches that define complex ideas. Searches are complicated by the need to address the constantly evolving vocabulary used to



describe clinical practice and technology. Keyword and full-text searching cannot compensate for this evolution in the terminology used to describe a field of research and often have limited mechanisms to understand all of the concepts embodied in a single term. In medical search, this is complicated by classification schemes built around medical coding requirements rather than a domain-specific lexicon.

### **Lack of Search Authority**

Searching for authoritative material in the Internet arena is complicated by the lack of a definitive authority on the integrity of the material presented by the search engine. Unlike peer-reviewed content found in medical journals, the techniques that cause search results to appear more quickly and higher in search results use page rank and hyperlink frequency as a qualifier. Searchers using conventional tools have no way to quickly determine the importance or authority of search results and often are faced with the daunting task of weeding out opinion from the most relevant current work in the target specialty or subspecialty.

Search engines (such as Google Scholar) and domain-specific permutations (such as Goldminer and Yottalook) use hyperlinks as a proxy for traditional citation analysis. The number of times a particular resource is “linked” to another paper results in the page’s rank increasing and the article being more frequently presented in searches. This approach is limited or ineffective given the types of complex searches that often arise in a diagnostic or clinical trial setting. Hyperlinks can be links to blogs, RSS feeds, and other opinion-based content that has not been vetted to assess scientific merit making search results untrustworthy. The investigator is forced to scan and review many non-pertinent results to find factual, current, and reliable results for use in decision making in patient care.

### **How It Works**

Our intelligent search tool retains the nuanced and conceptual framework created by the author and uses it to find other work that embodies the same concepts and ideas. The concept space index “understands” the content of the search source and compares it against its “understanding” of the content of the searchable database. Conceptual search understands complex queries ranking and returning results based on how close they are to the concepts in the search input.

The content of selected authoritative radiology journals is used to create the “biomedical imaging/radiological concept space,” a semantic framework defining the relationships between sets of documents and the terms they contain. This process produces a set of concepts that enables the search engine to “understand” the content. The semantic framework includes a term frequency and inverse document frequency array to support “up weighting” rare terms that are



important, reflecting the relative importance of the terms to the domain, and to address issues such as abbreviations and complex terminologies. This occurrence matrix contains the relationships among the terms and concepts. The concept space controls for problems found in other search techniques, including synonymy (the phenomenon in which different words describe the same idea; e.g., a “doctor” is also a “physician”) and polysemy (the phenomenon in which the same word has multiple meanings; e.g., spam – a meat product, and spam –unsolicited email) that result in return of irrelevant search results. The concept space knows the context of the term so that it can discriminate between the use of the word “tree” when used in a botany search versus a “tree” in the context of computer science. The concept space generator removes anecdotal instances of terms that can obfuscate the conceptual and contextual meaning of the content.

The IST’s latent semantic analysis of the source documents evaluates the contextual usage of the words by statistical computation that is applied to a large body of diagnostic imaging documents. The aggregate of all the word contexts in which a given word does or does not appear provides mutual constraints that determine the similarity of meaning of words and sets of words to each other. This creates word and passage meaning and closely mimics human word sorting and category judgment. The software extracts every contextual relationship among every term in every text object (document or other text source e.g. reports or textual device output) within the collection. This generates the vector space array representation of the term based on those relationships. This approximates the standard empirical growth functions of simple learning.

The fact that this compressive function begins anew with each context also yields a kind of spacing effect: the association of A and B will be greater if both appear in two different contexts than if they each appear twice in the same context. In a second transformation, each of these cell entries is divided by the entropy for the event type; the logarithmic measure of the rate of transfer of information in a particular message or language. These techniques make the association better represent the informative relation between the entities rather than the mere fact that they occurred together. The inverse entropy measure estimates the degree to which observing the occurrence of a component specifies what context it is in; the larger the entropy of, say, a word, the less information its observation transmits about the places it has occurred, so the less usage-defined meaning it has and, conversely, the less a particular contextual occurrence tells about its meaning.

Diagnostic imaging specialists provide guidance on the appropriate organization of the concept training set, which will define the concepts and terminology to keep the concept space relevant. The Coronado IST is further enhanced by integration of RadLex®, a free and easily accessible radiology lexicon created by subspecialty experts and organized by the Radiological Society of North America (RSNA). Users can use RadLex terminology and definitions to further refine their searches.



Once the conceptual space is defined, extensive collections of searchable documents from a variety of sources can be added to the searchable document set. The tool will incorporate the content of leading peer-reviewed research journals as part of the searchable document set. The searchable set will include all of the articles in the significant radiology journals and will be easily extensible to other journals or other documents, such as imaging reports, problem lists, or requests for imaging examinations. Abstracts archived in PubMed, a service of the U.S. National Library of Medicine that includes over 18 million citations from MEDLINE and other life science journals for biomedical articles dating back to 1948, will also be part of the searchable set.

### **Search The Way You Think**

This approach supports dynamic and fluid searching by allowing investigators to use research documents, definitions, emerging thoughts, and clinical best practice information as the basis of a search. Users simply copy and paste the text into the search box and search. Descriptions of diverse clinical issues, such as those uncovered during clinical trials or in complex patient diagnoses, can be combined to determine whether, for example, other practitioners have described a similar situation or protocol or to search for other researchers or specialists who have worked on the same problems. This approach removes the requirement for researchers to reframe their ideas or search topics into key terms delivering a more productive search experience.

There is no limit on the amount of content that can be used to a single execute a search. The more content used for a search the more refined the results. This allows users to explore a variety of strategies and approaches for finding the right answer to the question and the protocols to meet their needs. Coronado's intelligent search engine understands the changes in the terms used to describe technology and science over time, eliminating the need for researchers to determine how the concepts and discoveries in contemporary medicine might be described in older work and or how older techniques might be represented in contemporary research.

### **Search Outside Your Comfort Zone**

Conceptual search removes the requirement for researchers to understand the detailed lexicon of subspecialties to execute successful searches. Researchers don't need to understand the evolution of the language used to describe the practice and protocols in other specialties to find relevant information. This technology allows researchers to act upon their hunches, to formulate searches



that understand the concepts in the source and to find cross-discipline information that embodies the same concepts or ideas.

### **Advanced Single and Multi-Monitor Researcher's Desktops**

The conceptual search tool is accessed via an advanced researcher's desktop. The desktop user interface allows the researcher to create a project, to create and save searches, create and save search results, execute searches again to see if new research is available; enhance existing searches with new information. The interface supports a journal feature to allow researchers to keep their own notes or to create projects with other researchers to collaborate and share information. The desktop supports the domain ontology so that search criteria can be enhanced using its specific definitions and vocabulary to further refine and enhance search results.

Designed to support single and multi-monitor installations, researchers can create the electronic workspace needed to view and compare many resources at once supporting advanced analysis. The tool can run with other desktop tools allowing the user to cut and paste text based information from one application into the intelligent search tool to enhance search. If the users finds a resource that can enhance their research, the user copies and pastes the text into the concept search box and begins the search. Users can save both the search source and the search results or export them for use with other applications.

The user interface allows researchers to enter gigabytes of text-based articles, documents, text orders, diagnostic imaging reports, free form text disclosures, and other materials that define the complex search. A single search will search across all of the data sources indexed into the concept space returning articles and associated images that are closest conceptually to the search source. The intelligent search tool establishes result scores based on the relevance of the concepts in the results.

### **About Coronado Group, Ltd.**

Coronado Group, Ltd., is a specialized systems integration and technology consultancy focused on advanced search and document intense business processes. Coronado Group, Ltd. has used the same technology infrastructure defined here to build its Cognition IP product that supports conceptual searching of the entire US patent database.

Coronado Group designs very large-scale document imaging, document exploitation, and advanced analytical database systems. The firm has extensive experience building advanced analyst desktop and statistical quality control solutions for Commercial and Federal clients and provides advisory services to leading technology firms.



*Coronado Group, Ltd.*

*Intelligent Search for Answering Clinical Questions*

**For Additional Information Contact:**

Arleen Malley Zank

[arleen@coronadogroup.com](mailto:arleen@coronadogroup.com)

301-986-1334 ext. 109

703-909-7722 Mobile