Cs3, Inc.

5777 W. Century Blvd.
Suite 1185
Los Angeles, CA 90045-5600

Phone: 310-337-3013
Fax: 310-337-3012
E-mail: info@cs3-inc.com

Pioneering Technologies for a better Internet

TriggerWare™ ✦ Combo Search™

A Virtual Database Paradigm for Collaborative Internet Search Over Semi-Structured Data Sources

Dr. Donald Cohen
Dr. K. Narayanaswamy (**Point of Contact**)
Email: swamy@cs3-inc.com
Phone: (310) 337-3013

TriggerWare™ ✦ Combo Search™

Redefining Internet Search for Semi-Structured Data Sources

Breakthrough Search Technology:

Cs3 has built and is marketing ground-breaking, patent pending, search technology called **TriggerWare *Combo Search***. ***Combo Search*** is specifically intended to support new kinds of search operations for the part of the Internet that is "semi-structured" – which we define as data sources that have some useful structure that can be automatically retrieved through software.

Traditional search engines function by efficiently finding and displaying URLs that contain references to specific keywords. While this is the best one can do for web pages that contain no structure whatsoever, one can do a lot better for web-sites that *do* have internal structure that can be extracted through software. The ***Combo Search*** engine does not replace traditional search engines. Rather, the ***Combo Search*** engine complements traditional search engines. Traditional search engines will always be useful and necessary for exploring and browsing the Internet or learning about new areas. ***Combo Search*** enables much more focused, rapid use cases where the end point is the rendering of a decision. The TriggerWare technology accomplishes this objective by providing the following powerful, new features:

❖ **Viewing Web Pages as Tables:** Ability to view valuable data from various web pages as traditional tables. Tables provide an intuitive way for consumers to view structured data. **TriggerWare** allows any web page that has underlying structure to be viewed as a set of tables by the consumer.

❖ **Searching by Correlating Tables: TriggerWare** enables consumers to conduct "***combo searches***", whereby they can combine data from many different web pages in a single search. Note that a traditional search engine cannot do this in any meaningful way. ***Combo Search*** can compare and inter-relate information from several different web pages.

❖ **Results Displayed As Tables:** *Combo Search* displays results as tables, allowing consumers to render decisions faster and more conveniently. It also allows data providers to enable consumers to formulate useful connections between various kinds of data.

❖ **Making Search Queries Easy to Use:** *Combo Search* allows consumers to start using this powerful search technology without having to understand the underlying query language by creating a library of reusable combo searches. It is also possible for consumers to create new search queries on the fly if they choose to do so.

Concept of a Virtual Database

A simple and intuitive concept is needed to allow users to view the staggering variety of available data sources over the Web through a uniform prism. **TriggerWare** uses the notion of First-Order Logic *relations*, a generalization of the concept of relation or table in traditional relational databases as the common abstraction mechanism. Tables are something that average consumers intuitively understand. **TriggerWare** offers several core virtual database capabilities that prove to be crucial in data integration and Internet search problems. ***Combo Search*** is a specific application of the generic **TriggerWare** technology, whose core capabilities include**:**

Abstraction of the Concept of a Relation: The platform allows addition of new relational representations by supplying a set of interface methods (e.g., to test or generate the tuples of the relation under different conditions). Note that, unlike a traditional database, it is *not* required that all the tuples of a relation be fully generable.

Ability to View Many Computations as Relations: The interfaces above are generic enough that many computations (such as mathematical functions) can be viewed as relations. In addition, this flexibility allows a huge variety of data sources to be abstracted as relations (e.g., log files, web forms, and so on) in the virtual database.

Consistency Checking and Trigger Compilation: For transaction-oriented applications (note that search is NOT one of these), **TriggerWare** provide a transition language to describe data constraints, triggers over complex conditions in the virtual database, and several event correlation primitives. *Automatic* compilation of triggers written in the transition language supports the construction of applications that react rapidly and dynamically to real-world conditions that may affect key decisions.

Query Optimization: **TriggerWare** provides a query language over all the relations in the virtual database, independent of representation, thereby offering a powerful basis for search. The **TriggerWare** compiler is able to operationalize queries, constraints, and triggers, and can be guided to select more efficient algorithms using compile-time annotations supplied by the programmer. This feature supports performance tuning of an application.

How Does the TriggerWare✦Combo Search Solution Work?

Step 1: Analyze and Understand the Data Sources: The first step is to make sure that every data source of interest has *structure* that can be extracted through automated software. Data sources could be web pages, web forms, documents, and even other databases – *any* artifact from which one can **programatically** derive useful data.

Step 2: Define the Virtual Relation Metadata for Data Sources: The second step involves formally specifying the set of relations for each data source of interest. A data source can provide many different relations. Several working examples and building blocks of metadata specifications exist to guide new users on how to define the data sources properly. A unique feature of **TriggerWare** is that the metadata is potentially a *separate artifact* from the data it describes, and can be developed independently by somebody other than the data provider.
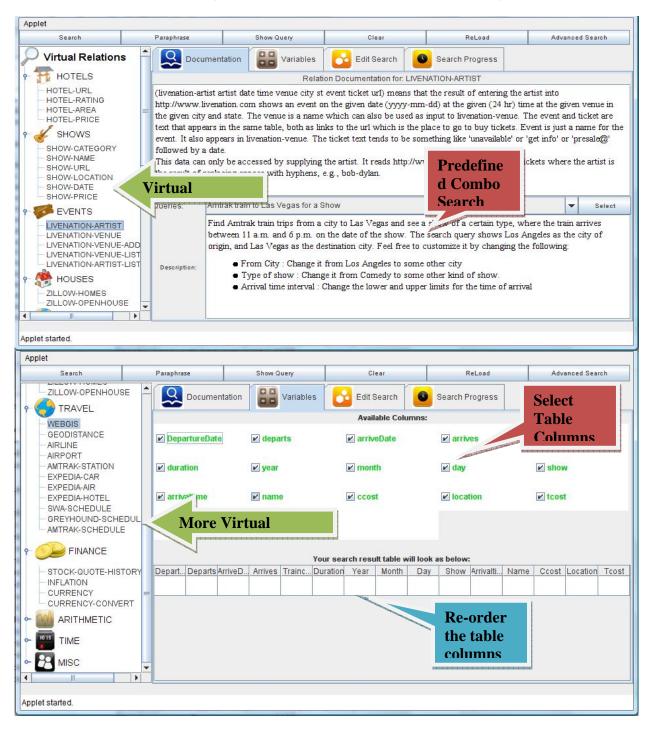
Step 3: Install the Metadata Definitions Into the TriggerWare Server: The metadata definitions need to be submitted to the **TriggerWare** server. The descriptions can be immediately verified and validated using simple *Combo Search* queries over the virtual relations that are submitted directly to the **TriggerWare** server.
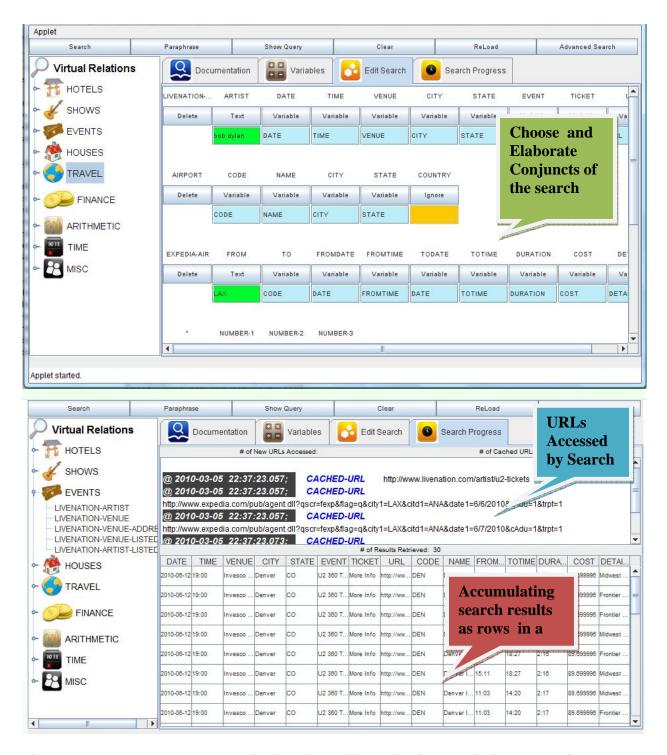
Step 4: Start Utilizing the Virtual Relations in Search Queries: As soon as the metadata specifications are validated in Step 3, those relations can participate in search with ALL the other relations in the virtual database. This is the power of this paradigm. In addition, if desired, **TriggerWare** provides a HTTP interface and direct sockets based interfaces to make the **TriggerWare** data available to application programs. In particular, applications that need asynchronous notification are supported by **TriggerWare** using the sockets-based API. Client programs can be written in a variety of languages, as convenient to the customer.

As described earlier, **TriggerWare** incorporates a query optimizer to select the most efficient algorithms possible for all search queries. The metadata specifier can guide the query optimizer by supplying annotations about expected data sizes and representations for virtual relations. In addition, because operationalizing a search query involves fetching data from real web pages over the Internet, network communications are an unavoidably large part of the cost. Data caching of web data to minimize network communication cost is, therefore, a critical part of the performance improvement strategy.

Current State of the Combo Search Technology

We include some screen shots from the *Combo Search* web client to provide an idea of some of the innovations relative to the traditional search experience. The **TriggerWare** server already incorporates metadata for many popular Internet sites already spanning travel, entertainment, real estate, financial, shopping, and other domains.

The **TriggerWare** engine is a mature technology that Cs3 has marketed in several other contexts for some time. The *Combo Search* web client, whose screen shots are shown here, is constantly evolving as we gain feedback on how users feel about the interface and the experience of creating searches.

Combo Search Technology Challenges

The major challenges that lie ahead for *Combo Search* include**:**

- **Extraction/Acquisition of Metadata:** Cs3 continuously seeks partnerships with companies that specialize in data extraction to scale up our metadata efforts in different domains. Metadata specification is not a trivial task. We have found that, while some aspects of the metadata specification process are amenable to automation, human supervision/input is still ideal to make sure the metadata description for a data source is appropriate and valid. We believe that the first company to create a representation of semi-structured data sources on the Internet can completely change what is possible with respect to Internet search.

- **Making it Easier for Consumers to Define Combo Searches:** We are constantly improving the GUI for advanced users to formulate their own searches. We currently provide drag and drop gestures to build up the query. The hardest problems have to do with specifying parameters – such as clarifying input columns, output columns, columns that are to be ignored and so forth. Finally, we are investigating how to guide users in gradually elaborating a search query through the use of an interactive Search Wizard.

- **Providing a Collaborative Foundation for Combo Search:** No matter how easy we make the GUI to define combo searches, average Internet consumers will not be able or willing to spend the time and energy to define new search queries in a formal language. To overcome this adoption barrier, each *Combo Search* is treated as an artifact that can be created, indexed, and published by one person, and reused by a much larger community of users. We intend to use Web 2.0 features on *Combo Search* to help users deploy *Combo Search* without the intellectual burden of having to define the queries themselves. In particular, we intend to create a Combo Search Widget to make it easier for users to use and distribute or share useful searches through all available channels, including social networks.

- **Integration with Customer/Partner Websites:** In order to make it easy for any website to integrate the *Combo Search* technology into their site, Cs3's current vision is to build an extensive library of reusable combo searches. Users could leverage the power of these queries without truly understanding the underlying query language. A REST-based API is available for websites to acquire the combo search options related to a given set of keywords, for example. Websites can integrate these options into their pages with minimal disruption to their current site and functionality.

Key Combo Search Technology Characteristics

TriggerWare server is the key technology component. The client side can be built in any number of languages. Please view the detailed data sheet for TriggerWare at http://www.cs3-inc.com/pubs/ps_TriggerWare.pdf .

How Customers/Partners Deploy Combo Search™

Combo Search is made available to customers of different profiles using business models that make sense for those customers. The business models include a modest initial flat fee for set up plus**:**

- ⊠ *Combo Search* as **a Service:** Website pays a reasonable, fixed recurring monthly fee to gain access to combo search options through a shared **TriggerWare** Server using a REST API.

- ⊠ *Combo Search* as **a Software Product:** Large websites can have their own **TriggerWare** server(s) to retain total flexibility to create new combo searches as needed for their customers.

- ⊠ **Revenue Sharing/Pay Per Click**: As **TriggerWare** gets deployed by customers and used in the marketplace, we intend to explore traffic-based and advertising-based business models **TriggerWare.**

Questions? Comments?

We welcome your comments (good, bad, or ugly)! We can also schedule online demonstrations of the technology, which is probably the best way to understand what the *Combo Search* technology can accomplish and how.