QL2 manageable data | on demand

Turning Unstructured Information into Actionable Market Intelligence – QL2 Technology & Solution Overview

WHITE PAPER

Situational Background

Business today is moving faster than ever. The internet has changed the shopping and purchasing habits of consumers in every sector. Despite the increasing volume of data available to guide marketing and business decisions, companies still rely on information that is not only too broad and expensive, it is often received too late to be useful.

While this type of information is publicly available on the internet, a large portion of it exists in unstructured form such as in web pages or online documents, rather than in traditional database formats. As a result, extracting the data and translating it into market intelligence presents a difficult challenge.

QL2 bridges the intelligence gap by providing the technology to automatically gather and present the most current, relevant, and comprehensive view of the marketplace. It empowers companies with the information to compete effectively, regardless of size or resources.

What is QL2 Technology?

QL2 technology is a flexible, intelligent technology solution that gathers structured and unstructured information from virtually any networked data source on the Web and within internal company sources. By collecting and categorizing specific information according to predefined rules, QL2 provides timely market intelligence that allows companies to make better decisions faster.

QL2 Use Case: Pharma Competitive Intelligence

A top-five global pharmaceutical company uses QL2 to collect broad types of market intelligence and to organize its internal content.

Specifically, QL2 collects and categorizes broad sets of market intelligence—encompassing public, protected (user ID and password) and private information including:

- Clinical trial status and adverse events information from FDA, CROs and clinical trial-specific websites.
- Formulary data from hospitals, medical groups and healthcare insurers.
- News directly from competitors and industry press.
- Detailed technical information from grants, research and conference proceedings.

QL2 aggregates purchased content and collected market intelligence into a centralized repository, allowing researchers to directly access this information through simple queries. A simple keyword search UI tool was built using QL2.

WebQL is QL2's proprietary programming language uniquely designed to collect and organize unstructured, uncontrolled data from websites and other complex network environments. WebQL can access information from virtually any document format, including data hidden behind forms, embedded in graphics, and contained within metadata. This requires sophisticated navigation and semi-structured document manipulation capabilities.

Designed for simplicity, WebQL does not require the installation of complicated hardware or software. In fact, minimal involvement from the IT department is needed to set up and run WebQL. With only a few lines of code written by a moderately skilled developer, WebQL can accomplish the same tasks as hundreds of lines of code written by a Java engineer. Also, since WebQL operates within a small technical footprint, it is unobtrusive to the data sources it accesses.

Much more efficient than web crawling, WebQL is highly scalable and capable of handling one query or thousands of queries with minimal CPU load. This seasoned and proven technology has made the customization and refinement of data extraction commands a simple task. In addition, WebQL streamlines the data acquisition process through simple mapping and lightweight normalization capabilities.

WebQL Compared to Other Technologies

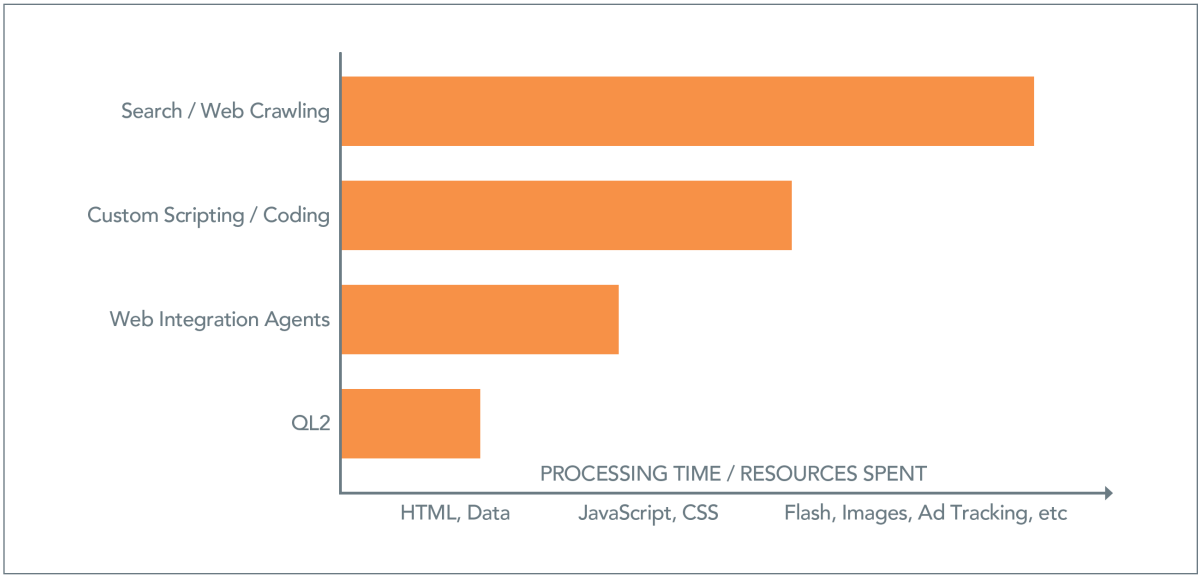

Diagram 1 – QL2's approach is dramatically faster than the alternatives since it leverages HTML and data elements that load on web pages first

When considering how to extract and organize information from the Web and other sources, technologists face many choices. Traditional messaging middleware as well as data integration and programming tools each offer a solution. However, these technologies face many limitations when compared to WebQL's ability to efficiently access unstructured information from uncontrolled data sources.

Technology	Limitations for Unstructured, Uncontrollable Information	QL2 Solution Benefit
ETL / EII	• Built for data warehousing • Often not possible for real-time information • Requires domain knowledge and developer skill sets across each source technology	• Agnostic to target source of data required • Designed to access information in real-time
EAI /Messaging Middleware	• Built for messages, not data collection • CPU intensive when performing large-scale data collection • Requires expensive developer skills	• Operates within a small technical footprint • No special developer skills required
Search / Web Crawling	• Built to collect everything, so it captures a lot of useless data • Cannot access protected and private information (behind forms, passwords, etc.)	• Collects targeted information • Unobtrusive to data sources • Designed to access public, protected and private information

Table 1 **(continued on next page)** – *Comparing Technologies*

Technology	Limitations for Unstructured, Uncontrollable Information	QL2 Solution Benefit
Web Integration / Agents	• GUI-based agents inefficient for targeted data extraction • Difficult to scale	• Built for enterprise scale • Intuitive development environment without GUI limitations
Scripting / Coding	• Custom approaches result in long development cycles and one-off solutions	• Automates manual scripting, reducing time dramatically

Table 1 **(continued from previous page)** – *Comparing Technologies*

QL2 Architecture

WebQL is the underpinning for our on demand services and is also available as packaged software for on-premise installation. Both deployment styles are built on the same underlying stack of technologies. QL2 technology is broken into four major areas:

- **WebQL**: Automates the collection and categorization of unstructured information.
- **QL2 Service Platform**: Transforms information into market intelligence.
- **QL2 Storage**: Stores the terabytes of data collected for QL2's on demand services.
- **QL2 On Demand Services**: Offered directly to business users to answer specific marketing or business questions.

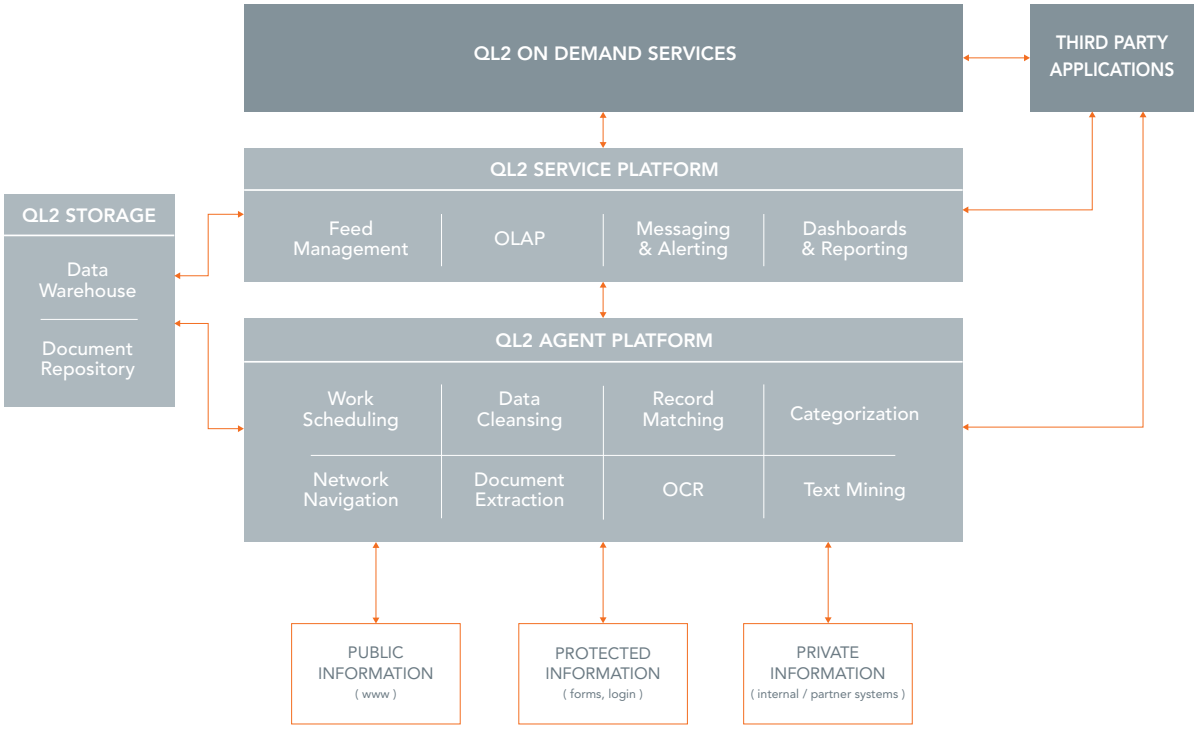

Diagram 2 – QL2 On Demand Services (Architecture)

As illustrated in the above diagram, WebQL collects information from a variety of sources including web-sites–both public and password protected–as well as file systems, data sources and enterprise applications within a client's firewall. The information gathered is organized and presented in QL2's on demand service or exported for integration into a wide range of external applications. QL2's Service and Storage Platforms manage the activities of data gathering agents and the storage of the data collected.

WebQL

WebQL is designed to automate information extraction by collecting and organizing public, protected, and private information. WebQL intelligently extracts information from any information source with features such as:

Scheduling: WebQL's scheduling feature can dictate when and how frequently the information is to be collected.

Network navigation: WebQL can navigate to any data that is accessible by a person through a web browser or other desktop application. WebQL handles all standard internet/web protocols (including HTTP/HTTPS for web data, ODBC for databases and many other specialized protocols such as FTP, IMAP, and several email protocols). WebQL can supply valid credentials to access information that requires authentication.

QL2 Use Case: Content Aggregation for Educational Publisher

A leading educational publisher uses QL2 to process millions of pages of PDF content from lesson planners provided to teachers as an addendum to textbook sales.

- Information varies by state (e.g., 3rd Grade Mathematics in California), and encompasses millions of pages of content in aggregate.
- Content is extracted, normalized and categorized to perform content verification and editorial tracking. This also provides a valuable audit trail.
- It is too expensive to gather data from the source –using QL2 is dramatically cheaper and faster.

Document extraction: WebQL understands dozens of common file formats, including HTML, XML, PDF, Microsoft Word and Excel, RTF, CSV, and many others. It is also able to pull data from compressed or archived formats such as GZIP and ZIP files.

Data extraction: A core feature of WebQL is its array of procedures for pinpointing data in documents ranging from extremely low-level processes like extraction-based regular expression patterns to high-level, heuristic techniques like identifying unstructured key and value data pairs.

Advanced Source Handling: WebQL can apply structure to whatever data format has been accessed, including CSV or Excel spreadsheets as well as tables in HTML or Word files. Using WebQL syntax, a developer can select all values in a table by simply naming the columns. Similar programming instructions can be used to extract data from many different native data formats.

PDF Handling: WebQL also has sophisticated PDF processing capabilities that can recognize table structures as well as font and style information.

Data Typing and Normalization: WebQL also has a sophisticated data typing and normalization mechanism that allows developers to specify constraints on floors and ceilings or advanced data types. This mechanism validates data at any point during execution:

- While reading input configuration information (to ensure that input parameters are valid).
- While generating intermediate data during processing (as a debugging tool).
- While generating output data (to ensure that the expectations of subsequent consumers of the data will be met).

OCR: Using Optical Character Recognition (OCR), WebQL is able to extract data that is only available as text embedded in an image, such as a product name or price. Because WebQL can read text represented as images in web pages and PDF documents, previously inaccessible data is now available for extraction and delivery.

The Service Platform

The Service Platform is where the information gathered by WebQL is transformed into market intelligence. Elements of this platform include:

Feed Management: Enables users to schedule and configure feeds–sets of information gathered by agents at regular intervals. This function provides users with a high degree of customization and automation when running regular information collection tasks.

QL2 Use Case: Order Application for Automobile Dealerships

An Automotive Solutions Provider developed a custom vehicle order application using QL2. QL2 delivered a dramatically simpler and cheaper application environment for the provider compared to traditional app-dev alternatives. This solution:

- Guides the user through complex manufacturer-specific vehicle configurations [make, model, color, options packages], collecting orders from hundreds of dealerships and submitting aggregated orders to manufacturers.
- Easily handles complex order configurations, including complex Java applets.
- Integrates with three major automobile manufacturers' systems.
- Is used by hundreds of automobile dealerships.

OLAP (online analytical processing): This element provides a structure for reporting the data. It pulls data from multiple sources and organizes it to provide a result.

Messaging and Alerting: The messaging function provides a concise summary of data gathering while the alerting function communicates behavior over a fixed period of time. For example, a client can elect to receive an email when certain thresholds regarding price and product availability are met.

Industry Dashboards and Reporting: The dashboard provides on-screen reporting of the market intelligence that WebQL accumulated over a long time span.

The Storage Platform

In an on demand world, storage is a critical component of any data gathering technology. QL2's Storage Platform includes the following elements:

Data Warehouse: Collects all extracted data in a secure, high redundancy environment for easy access.

Document Repository: Serves as a virtual warehouse that stores documents accessed during the data extraction process such as reports or website screen shots.

On Demand Services

QL2 On Demand services are hosted on QL2's On Demand platform, which accesses and analyzes over 500 million individual data points each month and supports more than 250 clients. The platform includes a variety of services, and allows for customization based on client needs. On Demand services include:

Pricing: Capturing pricing information from most websites consumers use to shop and delivering that information to clients on demand. This service provides QL2 clients with an instant snapshot of pricing and promotional activity across the competitive landscape. In addition, it enables clients to verify that their own pricing strategy is being properly implemented across channels and allows them to view competitor promotions, purchasing restrictions, and product availability.

Product: Allows clients to determine whether their category management strategies are working by comparing products and categories against those offered by the competition. Clients can view their product performance across multiple channels, conduct internal analysis, and identify market trends.

Custom Services: The flexibility of QL2 technology allows it to be easily customized to capture and organize any type of external market information from publicly available sources. There are several useful applications for this type of customization:

- **Compliance:** Check public data sources to verify whether the company is in compliance with local, state, and federal laws.
- **Procurement and Logistics:** Automatically shop for lowest prices on common goods and services to lower the company's overall operational costs.
- **Channel Management:** Ensure that downstream partners (distributors, resellers, retailers) are properly executing the company's programs.

On Premise Services

QL2 offers WebQL to clients who desire to maintain market intelligence on their own premises to maximize privacy and internal control. QL2's On Premise Solutions leverage the same platform as QL2's On Demand Solutions, and includes tools to create, automate, and analyze data collection and integration activity. QL2 technologies are readily deployed within a client's firewall and feature three levels of application:

QL2 Studio: The basic desktop product intended for clients who want to do their own in-house development using WebQL. QL2 Studio includes all the features and benefits of WebQL as well as a helpful debugging environment.

QL2 Server: Designed for large scale execution, QL2 Server is for clients engaged in high-volume, ongoing data integration activities. QL2 Server is scalable, allowing script execution to be coordinated concurrently across several machines, limited only by available hardware and network bandwidth.

QL2 SDK: Designed ebQL can be embedded in a client's own applications using QL2 SDK. For example, a client can develop an application that invokes WebQL and then returns the retrieved data to the developed application for subsequent processing. WebQL can be called from COM, Java or C++ application programming interfaces. Because WebQL has a very small memory and CPU footprint, it is easily embedded within a wide variety of applications.

Developer Efficiency

QL2 is a more efficient – and therefore more cost effective – development environment for extracting and organizing unstructured content.

- QL2 provides simple scripting and navigation tools for identifying and extracting content.
- QL2 is purpose-built to perform unstructured data extraction, automating tasks in a few lines of code that would require hundreds of lines of Java code.
- QL2 developers are less expensive to train than Java developers since it is a simpler task.

QL2 Solutions Turn Unstructured Information into Market Intelligence

By easily extracting and organizing data from virtually any website, network environment or document format, QL2 turns data into market intelligence, providing actionable metrics on what is being sold anywhere, at any time.

QL2 uses extensive HTML support that facilitates the collection of data points from unique online sources, using customer- and industry-specific pieces of software code known as agents. Because this technology is more efficient than Web crawling, it is scalable from one query to thousands with a minimal CPU load.

QL2 is available to clients both as an on demand service delivered in real-time through the Internet, and as on-premise software. In addition, it can be customized to meet specific company needs.