

Building your OSINT Capability

kapow

S O F T W A R E

White Paper

Kapow Katalyst for OSINT

Harvest text in any language, images, audio, video from websites, blogs, and social media.

Secure and non-attributable. Kapow Katalyst—*the best-kept secret in OSINT*.



HARVEST ANY OSINT DATA WITH KATALYST

OSINT data sources are as varied as the Internet itself. Mission-critical data can reside in blogs, in news feeds, in social media—and can even be hosted on short-lived sites on the dark web. As technology standards continue to evolve one thing is certain: OSINT data sources will remain a moving target—in more ways than one.

Only Kapow Katalyst™ has the deep understanding of Web page structure needed to handle any site, including those built with advanced AJAX techniques. Without that capability, data cannot be completely or accurately harvested from websites—and that's not good enough when national security is at stake.

CAPTURE DATA ANONYMOUSLY

Kapow Katalyst supports the widest range of options for non-attribution. Examine and alter JavaScript on any page before it is run. Automatically vary wait intervals, crawling schedule, and clicks per session. Alter HTTP headers to report different operating systems and browsers. Integrate with anonymization solutions. Alter proxy server in response to blocking attempts. Finally, don't draw attention to your mission by registering for API access—Katalyst can work with or without APIs.

EXTRACT DATA IN ANY LANGUAGE

Katalyst offers built-in support for multi-byte character encodings such as Chinese and bidirectional languages such as Arabic and Hebrew. Katalyst is in daily use throughout the IC to harvest the contents of news sites, blogs, RSS feeds, and social media around the world.

COPY AND RETAIN ANY WEBSITE—OFFLINE

Katalyst enables offline retention of any web page, domain, or list of domains for ongoing study and archival purposes. A Kapow SnapShot is a complete, fully static copy of any targeted site that includes all images, documents, and other linked assets. You can access it at any time without running the risk that the site owner will be alerted to your activity. Only Kapow offers both—a static SnapShot, and access to live websites.

ENFORCE ALL SECURITY REQUIREMENTS

Kapow Katalyst provides the widest range of security options—that's why it's in use on SIPRNET, JWICS, and other secure networks. Katalyst is stateless, so it can safely reside in your DMZ: there is no risk of compromising data because Katalyst retains no data. Contact us to discuss other deployable security measures.

PERFORM BROAD AND SURGICAL CRAWLS

OSINT projects use different crawl techniques at different stages of an investigation but most approaches and COTS products cannot accommodate this requirement. In the early stages a “broad” crawl is often used—but other solutions cannot scale sufficiently to support the high data volume that Katalyst can manage. Kapow Katalyst users routinely access thousands of web sites in a session with linear scaling. In later stages “surgical” crawls will seek to develop in-depth information on a narrow subject area. This type of crawl requires advanced website navigation and data transformation abilities and produces a much richer data set.

Surgical crawls can become quite sophisticated as analysis requirements become better understood. For example, it is often necessary to provide context and semantic richness as soon as an extracted item is recognized, rather than in a downstream process. For example, by recognizing a known terrorist during a broad crawl, a surgical crawl could be triggered to extract additional contextual data from internal and external sources. Then, by providing metadata and data facets based on context available only during the crawl—context unavailable later—downstream analysis tools can create timelines, scatter plots, heat maps, relationship graphs, and other advanced data representations to further assist analysis. Kapow customers routinely use Katalyst to perform this kind of access.

WORKING WITH KAPOW SOFTWARE

Kapow Software may be the best-kept secret in OSINT, but we’re easy to work with.

We’re GSA-listed and carry clearances.

We focus solely on OSINT Data Collection—and we’re good at it—so we work well with integrators and all the technology vendors in your OSINT environment.

We’re used daily throughout the IC to harvest data for Link Analysis and Visualization, Entity Extraction, Natural Language Processing, Search, Indexing.

For OSINT information, please call us at 703.489.1445 or email FederalSales@KapowSoftware.com

ACCESS THE DEEP WEB

Harvesting Deep Web data is much more complex than mere link-following, and Kapow Katalyst offers unique capabilities that enable Deep Web access. With its ability to navigate complex menu structures and generate input for web-based forms, Kapow Katalyst can automatically navigate arbitrary menu levels and issue multiple queries to extract entire databases and prepare them for use in any format necessary.

Data can be extracted from within content management systems and behind Web-based database query forms—then formatted for analyst use and automatically kept current with Katalyst’s de-duplication. Other extraction technology, limited to following specific links that already exist, are incapable of harvesting Deep Web data.

MAINTAIN MISSION READINESS

Kapow Katalyst uses patented technology that is resilient to page format changes, helping maintain readiness for any data capture your mission requires

ADD KAPOW TO YOUR OSINT TEAM

These are just a few of the capabilities that set Kapow Katalyst apart from other OSINT extraction technology. Contact us today to learn more about why we’re so widely deployed in the IC. Now and in the future, you’ll be pleased when you decide to make Kapow Katalyst part of your OSINT team.

ABOUT KAPOW SOFTWARE

Kapow Software is the leader in OSINT Data Extraction and Application Integration.

Kapow Katalyst™ automatically extracts, transforms, integrates and migrates data from virtually any source on the web, in the cloud or across the enterprise, to virtually anywhere, including mobile devices, without requiring APIs.

Nearly 500 customers, including AT&T, Audi, Intel, Morgan Stanley, Vodafone, and dozens of federal agencies rely on Kapow for data syndication, mobile-enablement, cloud and web app integration, content migration, business intelligence, terrorism tracking, comparison pricing – and much more.

Contents

Executive Summary.....	1
OSINT: Architecture Overview	2
Harvesting Data from the Internet	2
Modern Web Page Architecture	3
Data Harvesting for OSINT	4
Broad vs. Surgical Crawl	4
Extensible Data Extraction Support	4
Non-Attribution.....	5
Resilience to change	5
Flexible crawl strategy	5
Multiple language support.....	6
Security	6
Automated data de-duplication.....	7
Offline SnapShot	7
Advanced Data Normalization	7
Deep Web Access.....	7
Data Integration for OSINT	8
Broad-Based Support for Standards	8
Near Real-time Access	8
Partner Ecosystem	9
Mission Support for OSINT.....	10
Operational Agility	10
Scalability	10
Production Support.....	11
Build Your OSINT Capability with Kapow	12

Executive Summary

Intelligence agencies across the Federal government are turning to OSINT for the additional breadth and depth of information that it promises to deliver. For example, DOD Instruction 3115.12 directs DOD to "pursue the full integration and exploitation of open sources for intelligence purposes" and Intelligence Community Directive (ICD) 200 states that "Analysts must be able to find, access and share information, analytic judgments and each other in any media and in real time."

However, unlike typical IT systems which can accept only a limited range of input, OSINT data sources are as varied as the Internet itself—and will continue to evolve as technology standards evolve.

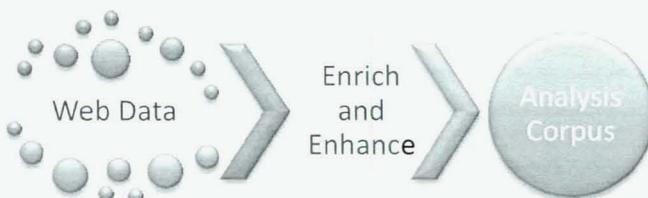
This white paper will alert you to the key issues that can derail any OSINT project. We will focus on the three major functional areas relevant to any OSINT capability:

- **Data Harvesting:** this is the “front end” of any OSINT process. Raw data is accessed, extracted from the irrelevant content that surrounds it, enriched with attribution that is available only as data is being harvested, and prepared for downstream use
- **Data Integration:** in this part of an OSINT process, raw data is integrated, enhanced, and enriched according to the requirements of the specific mission
- **Mission Support:** after an OSINT capability has been developed, additional functionality will be needed to support and enhance the capability as mission parameters change

This white paper will also illustrate the advantages that have led over 500 customers—including defense, intelligence, law enforcement, and other Federal agencies—to select Kapow as a key component of their OSINT data collection and acquisition strategies. With Kapow Katalyst these organizations discreetly capture more data faster, and respond more quickly to changing conditions.

OSINT: Architecture Overview

Just as each agency has a unique mission, each supporting OSINT initiative has unique requirements. However, after years of experience the OSINT community has begun to develop a set of best practices, coalescing around architecture broadly similar to the illustration on the right.



In this highly abstracted view, an initial step captures raw internet data and prepares it for a series of subsequent processing steps. These subsequent steps carry out a wide variety of mission-specific processing on the harvested data, enhancing and enriching it, and ultimately preparing a corpus for analysis. Finally, intelligence analysts access the corpus to carry out their work with additional mission-specific tools. In most cases a feedback loop is used when interim results expose the need to further augment the corpus by acquiring and enhancing additional raw data in a new cycle.

Harvesting Data from the Internet

Although the above abstraction applies to virtually all intelligence gathering and analysis efforts, OSINT practitioners focus on harnessing the breadth and depth of data available on the Internet—more specifically, on accessible websites both inside and outside their agency's firewall—and making this data meaningful and available to analysts. Extracting data from web pages, however, presents a large number of significant technical challenges that must be resolved before regular operational use can be made of Web-sourced data.



For example, extracting data from pages that use "Web 2.0" techniques—techniques that dynamically construct web page content and structure on-the-fly, as a page is being rendered—demands deep understanding of browser and JavaScript technology, and requires a protracted investment to ensure that evolving standards and browser technology are supported in a timely manner.

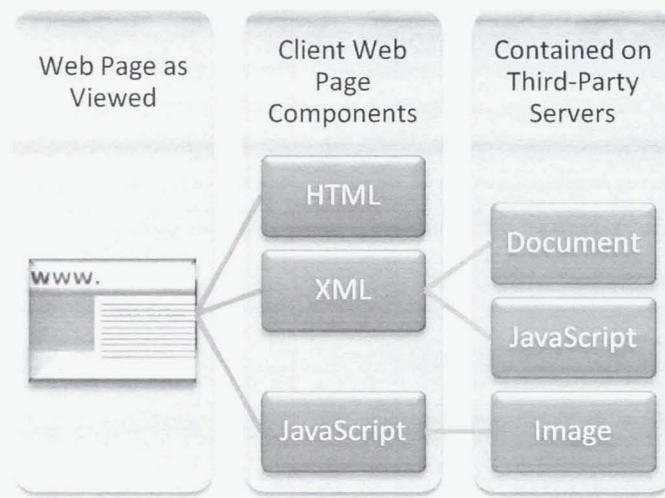
Kapow Software has the kind of deep understanding that can be earned only through a decade of experience, supported by continuous investment in browser-based data extraction. Kapow's unique Extraction Browser is in daily use across the defense and intelligence communities, extracting a wide variety of data from websites around the world and preparing it for further enrichment by members of our partner ecosystem.

Modern Web Page Architecture

The very complexity of a modern Web page forms a significant barrier to automated data extraction. Any given Web page is composed of a large number of elements with varying syntax and semantics—and because Web standards continue to evolve, continuous effort is needed to stay current.

Early Web pages were coded only with HTML, which was used only to specify how data was to be formatted for display. Each user request resulted in a complete unit of HTML being sent to the browser. However, as web standards have evolved, more advanced “dynamic” techniques have replaced static web page construction, presenting much greater technical challenges.

Dynamic techniques, known collectively as AJAX (Asynchronous JavaScript and XML), enable a page to be built under control of one or more JavaScript programs that run in the browser client as the page is loading. These JavaScript programs can perform a wide variety of actions—they can alter page format, request database access, send and receive messages, load media files or documents, and even cause other JavaScript programs to be retrieved and executed. In addition, JavaScript is commonly used to capture web page analytics and select user-specific advertising to be displayed along with the requested page content. Pages built with AJAX routinely involve multiple servers—not only the server hosting the selected Web page.



Some solutions use the Microsoft Windows HTTP Services (WinHTTP) API, but this has serious limitations. The API itself does not provide access to all of the page-creation activity necessary to correctly extract web data, and operationally it forces crawlers to run on Windows.

In contrast, Kapow’s Extraction Browser—used to support your crawlers as they run—contains a full JavaScript engine that enables Kapow Katalyst to see all dynamic activity as a web page is built, and Kapow also handles cookies, session data, authentication data, and other browsing artifacts that allow significantly more complete data extraction . One critical advantage that Kapow ‘s JavaScript engine provides is the ability for you to alter the flow on any web page in real time—an advantage that improves privacy and security as raw data is being captured. Finally, Kapow Katalyst supports Windows and several UNIX and Linux variations, giving you the widest range of deployment options.

Data Harvesting for OSINT

To better understand why Kapow has become the web data extraction tool of choice this section will examine each of the capabilities needed to harvest web data from a web site and prepare it for downstream enrichment and enhancement.

Broad vs. Surgical Crawl

Most OSINT efforts take the approach of casting a wide net during initial stages of investigation—often called a “broad crawl”—and focusing more sharply with a subsequent “surgical crawl” to fill important gaps in knowledge. Most automated approaches and COTS products fail because they are unable to accommodate both requirements.

A broad crawl will generate a high volume of data but most solutions cannot scale sufficiently to support high data volume. Kapow Katalyst is unique in this regard; it offers multi-server configurations that provide linear scalability capable of supporting the largest of crawls. Kapow customers routinely access thousands of web sites in a completely automated manner without experiencing bottlenecks.

Surgical crawls, often developed later in a project, typically require a different set of capabilities. For example, extracting data from news sites or blogs will require the ability to navigate through a series of menus, enter search criteria, scroll through a variable number of web pages, and extract only the latest data from a variable number of entries on each page. This enhances data quality and removes noise that diminishes the quality of subsequent analysis.

Surgical crawls can become quite sophisticated as analysis requirements become better understood. For example, it may be desirable to provide greater context and semantic richness for specific extracted items that are recognized, much as Google includes a map of Milton Keynes among the search results when “MK6 4HH” is entered. By recognizing a known terrorist during a broad crawl, a surgical crawl could be triggered to extract additional contextual data from internal and external sources. Then, by providing metadata and data facets based on context available only during the crawl—context which is not available later—downstream analysis tools can create timelines, scatter plots, heat maps, relationship graphs, and other advanced data representations that can further assist analysis. Kapow customers routinely use Katalyst to perform this kind of access.

Extensible Data Extraction Support

Because it’s not possible to predict what type of data will need to be extracted, both broad and surgical crawls will need the widest possible data extraction functionality. Katalyst includes the ability to extract any type of data from any Web site, including text, Office documents, PDF, flat files, images, audio and video files, and even files accessible via FTP or other non-HTTP protocols.

A wide range of transformation capabilities can be applied during initial data extraction or any time before data is delivered for further enrichment and enhancement, giving you the widest possible latitude in designing a flexible extraction and delivery strategy. As one example of the sophistication available, one customer that harvests academic papers from websites uses Katalyst to extract the

abstract from each PDF-formatted paper during the crawl—no other software is needed, and no code needed to be written. Many other similarly powerful functions are available for use, saving development time and improving crawl performance.

If you require specialized extraction or transformation capabilities beyond those provided with Kapow, you can invoke Java or .NET applications, as well as Web services, as you are crawling a site. Because you can create any specialized data extraction you wish, you can increase the quality of extracted data earlier in the process, improving responsiveness and moving closer to real-time extraction.

Non-Attribution

OSINT collection often requires repeated visits to a website, and it may become necessary to avoid being noticed. Kapow Katalyst supports the widest range of options for shielding your identity.

You can examine and alter the execution of any JavaScript code on a page before it is run, enabling a broad range of techniques for protecting identity such as defeating the use of JavaScript intended to track visitors. Kapow Katalyst also provides a wide range of ways to alter the flow of a crawler to make it more human-like—by inserting dynamic wait intervals, by crawling at different times of day, and by limiting the number of clicks in a crawl. Kapow Katalyst can also obscure its origin and appearance as it runs, by altering the HTTP header that contains identifying metadata such as operating system, browser/version, and referring URL. Kapow Katalyst can use proxy servers, can integrate with third-party anonymization solutions, and can even detect when access has been blocked during a crawl and respond by changing proxy servers dynamically during a crawl. Finally, by enabling access to a site's UI, Kapow Katalyst lets you avoid registering for API access—likely drawing less attention.

Resilience to change

Web sites change daily and some data extraction techniques are highly sensitive to even the smallest changes in page layout. Kapow's patented Extraction Browser is specifically designed to be resilient.

The Extraction Browser has deep understanding of Web page structure, and operates on this underlying structure of a page. Parameter-driven “tag finder” definitions mean that custom tags and syntax errors do not disrupt processing, and help shield the Extraction Browser from future changes to standards and technology.

Flexible crawl strategy

Each OSINT crawler will need to create a crawl strategy tailored to the specifics of the data sources and mission requirements, and strategies will vary widely. The Kapow Extraction Browser provides the widest range of options for designing a crawl.

Kapow Katalyst can automatically authenticate to a site, perform complex navigation through menu structures, supply input for any web form, navigate through multiple pages of output, and scroll through multi-screen web pages extracting multiple occurrences of data. Crawlers can test for the presence or absence of data or metadata anywhere on a page, and can change the course of a crawl at any point in time. Extraction rules can process as many paths for a page as desired.

In one case Kapow Catalyst was able to penetrate a Flash menu built expressly to protect a hacker site from “web scraper” tools, which simply do not offer the same functionality

Multiple language support

Whenever multiple language support is a requirement, Kapow Catalyst offers built-in support for single- and multi-byte character encodings, as well as support for bidirectional languages such as Arabic and Hebrew. Kapow Catalyst correctly captures and processes data from news sites, blogs, RSS feeds, and social media sites around the world.

The screen image to the right illustrates Kapow’s Design Studio being used to extract data from an Arabic language website. The Design Studio’s graphical editor allows precise visualization of selected data in any language and encoding, enabling point-and-click extraction of data from any page—even when the website does not advertise the encoding.



As a pure Java application, Kapow Catalyst uses Unicode natively, and data can be saved in UTF-8 or UCS-16. The Kapow Software product line follows all I18n guidelines, and multi-byte support is fully integrated throughout.

Security

Many websites require a user to authenticate before accessing some or all of the site’s contents. The Extraction Browser has the ability to enter any user ID and password into the site’s authentication page, just as it can mimic any other action a human user would perform (such as navigating through menus, scrolling through multiple page outputs, and entering data into forms). These access credentials can be stored for automated use, or a user can be required to re-enter them for each crawl, according to requirements. If a site uses PKI (Public Key Infrastructure), Kapow also has the ability to use authenticate using a digital certificate.

Several additional security measures can be applied. You can restrict data transfer to a specific IP address, port, and socket, and can use SSL for all communication among Kapow Software components. API access to Kapow crawlers can support ACL and security, and additional security measures can be discussed on request. Finally, the Kapow Extraction Browser is stateless, so it can safely reside in your DMZ: there is no risk of compromising data because no data is retained by the Extraction Browser.

Kapow Katalyst is currently in use on several secure networks, including SIPRNET, JWICS, and NMCI.

Automated data de-duplication

Only a small fraction of a typical website's content ever changes once it has been placed online. On a news site, for example, even though the relative newsworthiness of an article will alter the article's position on the site over time, the article itself may never be updated—and the Extraction Browser can detect whether the article has actually changed. Kapow Katalyst can report downstream whenever data has been added to a site, changed, or removed from a site, and duplicate data will never be delivered unless requested.

Offline SnapShot

Studying a private copy of a site—instead of the live site itself—offers a number of advantages, so Kapow provides the capability to take a “SnapShot” of any web page, domain, or list of domains.

A Kapow SnapShot is a complete, fully static copy created by extracting the results of all dynamic page construction activity. A Kapow SnapShot includes all images, documents, and other linked assets; all links are automatically rewritten to the local server where the SnapShot has been stored, and all accompanying metadata is written to XML or into a database for later analysis.

You can access a SnapShot offline without running the risk that the site owner will be alerted to your interest due to JavaScript execution of tracking code on the live page. Only Kapow offers both—a static SnapShot, and the ability to harvest data from a live site.

Advanced Data Normalization

Common on websites is the use of relative dates and time stamps such as “one minute ago” or “yesterday” and the extraction technology must be able to convert these into usable data. Kapow’s ability to convert relative times into usable date/time stamps is especially important to engage at the point of capture, because this data cannot be accurately recreated downstream.

Kapow can also access metadata and expose it as data usable by downstream processes. For example, some sites use color to indicate the status of an item on the page; other solutions cannot normalize this type of metadata into data that is useful.

Deep Web Access

Not all data is readily available on standard web pages. Experts estimate that the “Deep Web”—consisting of web pages built from queried databases and similar dynamic content—may be 20x larger than the “Surface Web” that is accessible by crawling hyperlinks. OSINT programs cannot afford to ignore the Deep Web, but harvesting Deep Web data is much more complex than mere link-following.

For example, a given mission may require capturing data from the databases and content management systems used to support a particular website, and storing it in a local database designed for analyst use. With its ability to navigate complex menu structures and generate input for web-based forms, Kapow Katalyst can automatically issue multiple queries to display the desired data. Coupled with its ability to navigate through multiple pages of output and extract data from multiple items on each page, Kapow

Katalyst can readily extract and download entire databases and prepare them for use in any format necessary. Kapow's de-duplication can be used to automatically keep it current. Other technology, limited to following specific links that already exist, will be unable to perform.

Data Integration for OSINT

This section examines requirements generated by the need to integrate raw data into downstream enrichment and enhancement processes.

Broad-Based Support for Standards

A recent Federal best practice study makes it clear that standards-based products are essential for any successful OSINT initiative. Because no single vendor provides everything a mission will require, it will be necessary for each OSINT program to integrate the technology it requires—and this integration will be much easier when proprietary solutions are avoided. Processes that provide raw data to downstream enhancement and enrichment technology occupy a special position in the architecture, and bear a concomitant responsibility to support the widest possible range of industry standards; otherwise, implementation options will be needlessly restricted.

Kapow supports the widest range of integration standards available. In addition to native support for XML, HTML, JSON, JMS, and CSV, Kapow can also deliver results into SQL databases and XML databases—and can create numerous other formats. Data can be exposed as Web services (SOAP or RESTful) and can be callable via Java or .NET interfaces. Unlike other solutions, Kapow Katalyst-based crawlers can direct output to any of these formats without being updated—a run-time choice selects the desired output format.

Finally, Kapow Katalyst can drive content into any Web application via its UI, by automatically completing the forms used to control the application. No other solution provides this range of design alternatives for planning your OSINT architecture.

Near Real-time Access

The same Federal best practice study also points out the desirability of “enriching the dataset as it is crawled” to enable near-real time access and improve responsiveness—especially important when conditions are changing rapidly. This can be accomplished by having the crawler manage some portion of the enrichment (by invoking whatever enrichment services are needed) before it pushes the enriched data downstream (either directly into the corpus or via whatever standards-based interface has been chosen).

Just as Kapow can invoke Web services, Java, or .NET applications during a crawl, it is possible to configure Kapow to directly invoke downstream enrichment capabilities during a crawl. This capability (perhaps most useful during a surgical crawl where a small amount of data needs enhancement) allows the OSINT architect to decide based on mission parameters, and not technical limitations, whether real-time enhancement is needed.

Partner Ecosystem

Although any successful OSINT effort will combine a variety of common techniques such as geotagging, entity extraction, language translation, and other types of enrichment, each OSINT initiative is unique. Depending on its mission, each project will be a mix of assembling COTS software and creating mission-specific capabilities, but cost and risk will be reduced by choosing technology that has already been proven to interoperate.

Kapow Software has an extensive partner ecosystem that includes key vendors and integrators that specialize in Federal projects, and we have jointly delivered a variety of OSINT projects. You can feel comfortable focusing on your OSINT requirements, knowing that Kapow and its partners are dedicated to your success.

Mission Support for OSINT

This final section discusses issues relevant to launching and operating your OSINT capability.

Operational Agility

Most OSINT missions will need to change over time: sometimes slowly, as mission parameters are refined through greater understanding, and sometimes rapidly, in response to conditions on the ground. Crawlers will need to be created or enhanced as greater understanding is attained, and new types of data will need to be extracted as they are found to be relevant to the mission.

Kapow provides agility in multiple ways. First, unlike other solutions, Kapow does not require a separate crawler for each site. Sites that share similar characteristics—blogs, for example—can typically use the same crawler because Kapow operates on the underlying Web page structure, not the presentation details of a given site. Kapow Catalyst also enables run-time redirection of output to a different format (e.g. SQL database, XML, CSV, RSS) allowing a given crawler to meet the needs of multiple downstream processes over time. Any crawler can be exposed as a SOAP or RESTful service, or called from .NET or Java; when combined with Kapow’s resilience to web site changes this gives the best kind of agility, where an external change incurs zero, or near-zero cost.

When it comes time to extract additional raw data—say, to develop a new surgical crawl—Kapow’s integrated repository makes it easy to share extraction and transformation rules, and the Design Studio supports rapid, point-and click creation of new crawlers, new extraction rules, and new transformation rules—often in hours. Analysts can develop their own automated data collection processes, and run them on demand or on a scheduled basis. Kapow also lends itself to a Center of Excellence organization if preferred, allowing extraction skills to be collected in a shared-services model.

Finally, Kapow provides the support you need to handle real-time enhancements to mission requirements as conditions change on the ground. For example, in response to the January 2011 government-initiated Internet blocking in Egypt, Google created www.speak2tweet.com—a service that recorded voice messages left on a pool of standard phone lines and automatically created a Twitter post with a link to the recording. Enhancing a Kapow crawler to extract the audio and pass it downstream for storage and analysis would take under 15 minutes, start to finish.

Scalability

Performance is critical in any large-scale initiative such as OSINT data capture, and high performance needs to be designed in from the start. Kapow Catalyst was designed with linear scalability in mind, achieved through advanced design and rigorous testing.

Kapow’s stateless design improves security, scalability, and availability, and supports the use of internal cloud-like resources to enable smooth scaling upward during periods of high load. Each crawl runs independently, in its own thread, providing the linear scalability needed to support extremely high volume crawls.

Production Support

For OSINT systems that will be run on a production basis it will be necessary to have access to a high level of support, both from the vendor as well as from internal IT resources. Kapow provides secure web-based tools that simplify all phases of IT involvement.

Kapow offers user, role, and group-based control over administration and deployment of the crawler environment, allowing secure deployment even in a shared-services organization. Fully-automatic scheduled crawls are provided, along with support for major monitoring standards such as JMX to ensure smooth execution and rapid detection of outages should they occur. Kapow Catalyst provides error logging that can be directed to files, databases, emails, and even custom formats—all completely secured. Error correction and redeployment are integrated, requiring just a click to restart.

Finally, as a company Kapow's support team is available 24x7 to assist with software issues.

Build Your OSINT Capability with Kapow

Kapow's Extraction Browser is a perfect match for the needs of OSINT.

Kapow can fully automate data extraction from any Web site. Without human intervention, Kapow Katalyst can authenticate, perform complex navigation, carry out sophisticated extraction and transformation rules, and reliably deliver data wherever it is needed.

Kapow Katalyst is fully standards-based, allowing smooth exchange of data in any required form. Data can also be exchanged via Web Services, Java, or .NET functions.

Kapow's integrated environment allows you to build even the most sophisticated crawlers without the need to develop any code, and test them in real-time—increasing agility and shrinking effort to a minimum.

An integrated scheduler allows even high-volume crawls to be run on any schedule, and provides complete production-level support to enhance availability.

Kapow—and our extensive network of technology vendors and integrators—are ready to serve you.

Contact us today, and join the other OSINT projects that are benefiting from Kapow.

Kapow Software

260 Sheridan Ave, Suite 420
Palo Alto, CA 94306
Phone: +1 800 805 0828
Fax: +1 650 330 1062

Email: Sales@KapowSoftware.com
Website: KapowSoftware.com
Blog: KapowSoftware.com/blog
Twitter: @KapowSoftware

107-11 Fleet Street
London, EC4A 2AB
Phone: +44 207 936 9266