CIQ Technology Services Specifications

SearchDocument

Version: 1.0

Date Created: 6.5.2008

Last Updated: 4.3.2009

Business Owner:	Jay Zachter, Michael Yusko
Technology Owner:	William Murphy (CIQ),
Business Analysts	Shawn West
Version	1.0

Scenario

This document represents the specifications for SearchDocument. It allows the client application to search specific document types stored in Capital IQ. These Types include filings from the Securities Exchange Commission (SEC), System for Electronic Document Analysis and Retrieval (SEDAR), Australian Stock Exchange (ASX) and Annual Reports. Clients may be required to purchase a separate license for the use of these data sets. Contact your Client Development Representative for details.

Application Framework

The primary technology for this solution is XML Web Services (SOAP). Capital IQ hosts an API that responds to XML requests according to this API, and returns XML structured data in response. These XML requests are encrypted via the standard HTTPS protocol.

A secondary technology for this solution is the integration of CIQ DataFeeds on client database tier. This allows for reduced network traffic for common items that change infrequently.

Capital IQ hosts this data on Windows-based servers, powered by Microsoft SQL Server in an active-passive failover cluster configuration. Data is stored in multiple fully redundant EMC Storage Area Networks (SANs). The servers that run the platform are hosted at Quality Technology Services with a disaster recovery site at XO. At all levels, these environments are redundant, fault tolerant, and backed up to industry standards.

Web Services Description Language (WSDL) documents describe the detailed Services & Ports (Function Calls) available in this specification. See http://www.w3.org/TR/wsdl for more on WSDL.

Please note that all Web Service and WSDL URLs in this document are subject to change based on changing infrastructure requirements. CIQ will provide sufficient advanced notice to the client before changing any URL, hostname, IP address, etc. It is recommended that these URLs be configurable (via config files, etc.) on the client application so that changes can be handled with minimal user downtime. CIQ monitors activity on Production systems and may shut down improper-use processes or user accounts as required to preserve overall system health.

All Web Services requests and responses in this solution are encoded in the UTF-8 character set (http://en.wikipedia.org/wiki/UTF-8). Some string data in this solution is expected to only contain Windows-1252 characters (http://en.wikipedia.org/wiki/Windows-1252); these are labeled with "(W1252)" in this document. Other string data in this solution allows full UTF-8 characters; these are labeled with "(UTF-8)" in this document. Email addresses (labeled "(email)" in this document) and website URLs (labeled "(URL)" in this document) have more limited valid character sets. See http://en.wikipedia.org/wiki/Email_address and http://en.wikipedia.org/wiki/URL for more information.

All the web services have a WSDL definition that external developers will code against and pull in data that is served from the same Capital IQ data repository as our web platform. For a full menu of our Web Services and implementation documentation, please contact your account manager.

Web Service Versioning

Versioning Web Services: Over time, Capital IQ may need to extend the tags or datasets supported by our Web services. As a results we have created a URL based versioning solution provides a scalable framework for the future. Versioning provides a way for to accommodate these enhancements in a graceful manner.

Recommendation: Capital IQ recommends that all users upgrade to version 1.0 if they are using legacy services, to conform to the new URL formats.

How versioning works: Please note in the example below *<ServiceName.asmx>* is replaced with the name of the service and is used for illustration purposes only.

i. Web Service changes are captured as a new version of the file in a new directory.

   a. **Version 1** - https://api.capitaliq.com/ciqdotnet/api/1.0/*<ServiceName.asmx>* - Represents the first release of the service
   b. **Version 2** - https://api.capitaliq.com/ciqdotnet/api/2.0/*<ServiceName.asmx>* - Represents the second release and breaking change or significant enhancement.
   c. Clients have the ability to transition to the new version of the service or stay on the original version until they can transition older code.

ii. Latest version of the Service will be located at the following URL. https://api.capitaliq.com/ciqdotnet/api/current/*<ServiceName.asmx>*. Using the example in section i above https://api.capitaliq.com/ciqdotnet/api/2.0/*<ServiceName.asmx>* would be in its own directory and referenced in the current directory.

Service Changes

Search Documents

Service	Version	Comments
URL	Current	https://api.capitaliq.com/ciqdotnet/api/Current/SearchDocuments.asmx?WSDL
URL	1.0	https://api.capitaliq.com/CIQDotNet/api/1.0/SearchDocuments.asmx?WSDL

Release	Version	Comments
9/2008	1.0/Current	Updated to conform to Capital IQ new versioning criteria

Documents Search Summary

Documents Summary

This service returns information about filings for a given Company ID, optionally limited by date, form types, and keywords. The document information returned includes SEC, SEDAR, ASX and several other sources.

The URL for downloading the documents is returned by the get download service, but can also be built manually as follows:

1. URI-stem /CIQDotNet/Filings/DocumentRedirector.axd
2. versionId versiondId returned in SearchDocuments
3. type output format type string returned in SearchDocuments

In some cases documents are not available on a real time basis according to contractual constraints. If a document is requested prior to its availability an HTTP error 403 Forbidden will be returned.

Documents Ports (Functions):

```
DocumentInfo() SearchDocuments(
               Integer CompanyId(),
               Datetime FilingStartDateUTC, Datetime FilingEndDateUTC,
               Datetime ProcessedStartDateUTC, Datetime ProcessedEndDateUTC,
               Integer SourceId (), Integer FormTypeId(), Integer FormTypeGroupId(),
               String CompanyName(), String  KeywordFullText);
```

Comments:

Client Note: For SEC Documents the SupplierDocumentId type is Accession Number, a unique identifier issued by the SEC for each filing.

Client Note: The number of documents returned by the service is restricted to a maximum count for performance considerations. This limit is subject to change.

Client Note: Reference data listed in the descriptions in this document may not be complete and may change over time. Please retrieve the latest reference data using the ReferenceData web service and the ID mentioned in the developer notes below.

Parameters:

4. `Array of Integer` **CompanyId** – The list of Capital IQ company identifiers whose documents will be returned. **Input** [Optional], [Multiple]

5. `Datetime` **FilingStartDateUTC** – Date in the past, defines the starting point for the document request. The filing date is based on when the company states the filing is submitted. This may differ from the actual date & time the document was publicly disseminated for a number of reasons, e.g. no time provided, corrections, pre-publishing, historical processing by Capital IQ, etc. **Input** [Required if ProcessedStartDate not populated], [Single]

6. `Datetime` **FilingEndDateUTC** – Defines the end point for the document request. If not populated the service will return all documents filed after the FilingStartDate. **Input** [Optional], [Single]

7. `Datetime` **ProcessedStartDateUTC** – Defines the starting point for the document request. This is based on the actual time the document was processed by Capital IQ. **Input** [Required if FilingStartDate not populated], [Single]

8. `Datetime` **ProcessedEndDateUTC** – Defines the end point for the document request. This is based on the actual time the document was processed by Capital IQ. **Input** [Optional], [Single]

9. `Array of Integer` **SourceId** – List of document sources to retrieve, this could be SEC, SEDAR, ASX, etc.
   Input [Optional], [Multiple].

   Developer Note: Refer to Reference Data ID 95.

ID	Name
0	SEC
7	SEDAR
16	ASX

10. `Array of Integer` **FormTypeId** – Only use with SEC Documents for the Form Type returned in the list.
    Input [Optional] [Multiple] possible values.

    Developer Note: Refer to Reference Data ID 28.

ID	Name
1	10-12B
2	10-12B/A
3	10-12G
4	10-12G/A
5	10-C
6	10-C
6	10-C/A
7	10-K
8	10-K/A
…	etc

11. `Array of Integer` **FormTypeGroupId** – Only use with SEC Documents for the Form Type Group returned in the list.
    Input [Optional], [Multiple] Possible values:

    Developer Note: Refer to Reference Data ID 94.

ID	Name
1	Annual Reports (10-K, 20-F)
2	Quarterly Reports - 10Q
3	Proxies (DEF, PRE)
4	Institutional and Mutual Fund Ownership (13F)
5	Insider Holdings (3, 4, 5, 144, 13D, 13G)
6	Others
7	Registration and Prospectuses (S-1, SB-2, 424)
8	Termination Delistings and Suspensions (Form 15)
9	Tender Offer Notices (SC TO, SC14D9C)
10	Corporate Releases and Announcements (8-K, 6-K)
11	All Forms Except Insider Holdings

12. `Array of String` **CompanyName** – List of company names to search

13. `String` **KeywordFullText** – Text to find in the full content within documents.

Returns:

Client Note: The sort order for the returned documents is always Filing Date descending (i.e. most recent first).

1. An Array of **DocumentInfo** Objects – Each contains information about one Document.

DocumentInfo ()
Attributes:

a. Array of FilingEntity **Companies** – List of companies associated with the filing. This is usually the filer, but in the case of ownership filings, this will be the issuer of the holding.

b. Array of FilingEntity **ReportingOwners** – Used primarily for ownership filings, this is a list of the people or companies reporting their ownership positions in the filing.

c. Integer **VersionId** – Some "documents" contain multiple "resources", "formats" or "revisions". CIQ combines these concepts into "Version", which is uniquely identifiable by versionID. Multiple Versions can exist for the same Document. For example, in a filing there may be two versionIDs: one for the original file and another for a revised corrected version.

d. Integer **DocumentId** – This is an identifier that is unique across documents (research, SEC, etc.) A Document is unique set of content independent of formatting, meta data, etc, e.g Apple's 10-K report

e. String **SupplierDocumentId** – A unique identifier for a document, issued by the source, i.e. SEC accession number.

f. Integer **FormTypeId** – The Capital IQ form type identifier. This identifier is unique across all sources

g. Integer **Description** –The textual representation of the Form Type or Headline.

h. Datetime **FiledDate** – Date and time when the document is filed with the exchange or regulatory body.

i. Datetime **PeriodEndDate** – The end date of the relative financial period for the document

j. Datetime **ProcessedDate** – Date and time CIQ processed the document

k. Array of DocumentOutputType **VersionFormatTypeIds** – List of possible output formats that can be requested for this document.

DocumentOutputType
html
xls
doc
pdf
raw

l. Integer **FileSizeInBytes** – The size of the file, in bytes.
m. Integer **SourceId** – The provider of the filing, i.e. SEC, SEDAR, ASX, etc.

Exceptions:

1. An exception will be thrown if the request cannot be authenticated via a session cookie.
2. An exception will be thrown if any parameter is out of range.

Web Services Description Language (WSDL)

Notwithstanding anything to the contrary in this Agreement, Capital IQ reserves the right to change, expand or modify Web Services Definitions and corresponding Web Services Description Language files (WSDL) at any time. Any such modifications are in accordance with industry standards that support backwards compatibility with previous WSDL files.

Client Note: Login using the supplied UserName and Password provided by Client Support or your Client Development representative.

Appendices

1. **Windows-1252 A character encoding of the Latin alphabet**, used by default in the legacy components of Microsoft Windows in English and some other Western languages. The encoding is a superset of ISO 8859-1, but differs from the IANA's ISO-8859-1 by using displayable characters rather than control characters in the 0x80 to 0x9F range. It is known to Windows by the code page number 1252, and by the IANA-approved name "windows-1252". This code page also contains all the printable characters that are in ISO 8859-15 (though some are mapped to different code points).

2. **Extensible Markup Language (XML)** is a general-purpose markup language. Its primary purpose is to facilitate the sharing of data across different information systems, particularly via the Internet.

3. **dateTime [Definition:]** values may be viewed as objects with integer-valued year, month, day, hour and minute properties, a decimal-valued second property, and a Boolean timezoned property. Each such object also has one decimal-valued method or computed property, timeOnTimeline, whose value is always a decimal number; the values are dimensioned in seconds, the integer 0 is 0001-01-01T00:00:00 and the value of timeOnTimeline for other dateTime values is computed using the Gregorian algorithm as modified for leap-seconds. The timeOnTimeline values form two related "timelines", one for timezoned values and one for non-timezoned values. Each timeline is a copy of the ·value space· of decimal, with integers given units of seconds.

   The ·value space· of dateTime is closely related to the dates and times described in ISO 8601. For clarity, the text above specifies a particular origin point for the timeline. It should be noted, however, that schema processors need not expose the timeOnTimeline value to schema users, and there is no requirement that a timeline-based implementation use the particular origin described here in its internal representation. Other interpretations of the ·value space· which lead to the same results (i.e., are isomorphic) are of course acceptable.

   All timezoned times are Coordinated Universal Time (UTC, sometimes called "Greenwich Mean Time"). Other timezones indicated in lexical representations are converted to UTC during conversion of literals to values. "Local" or untimezoned times are presumed to be the time in the timezone of some unspecified locality as prescribed by the appropriate legal authority; currently there are no legally prescribed timezones which are durations whose magnitude is greater than 14 hours. The value of each numeric-valued property (other than timeOnTimeline) is limited to the maximum value within the interval determined by the next-higher property. For example, the day value can never be 32, and cannot even be 29 for month 02 and year 2002 (February 2002). For more details http://www.w3.org/TR/xmlschema-2/#dateTime