



MOSS 2007 Open Calais Integration

Version: 1.1

August 2008

Table of Contents

1. Introduction.....	1
1.1. Intelligent Search	1
1.2. Open Calais Semantic Analysis.....	1
1.3. Rich Content on Your Web Site	1
2. How It Works	3
2.1. Example	3
3. Using MOSS 2007 Open Calais Integration.....	5
3.1. License.....	5
3.2. System Requirements	5
3.3. Installation.....	5
3.4. Configuration.....	5
3.5. Relative Tag Base URL	7
3.6. Activation	7
3.7. Verification	8
3.8. Troubleshooting.....	8

1. Introduction

The MOSS 2007 Open Calais Integration automatically adds semantic metadata and intelligent search capabilities to content pages on Microsoft Office SharePoint Server 2007 (MOSS 2007) content management web sites. This document provides an overview of the integration and information on installing and using it.

Basic knowledge about the MOSS 2007 infrastructure is required; this document assumes you are familiar with the terminology of MOSS 2007 and its publishing infrastructure.

1.1. Intelligent Search

Search engines allow filtering documents based on keywords that appear in a web site. One word can often refer to several different things. For example, when you search for “Washington” using a search engine, results will likely relate to the city, the state and the person.

Intelligent search allows users to specify with greater precision what they are looking for. For example, a user may ask for results relating to “Washington,” but only on reference to the state.

1.2. Open Calais Semantic Analysis

The Open Calais Web Service can analyze text and provide rich semantic data to provided text. For example, when the word “Washington” appears in the text, the Open Calais Web Service can determine whether it refers to the city, the state or the person.

1.3. Rich Content on Your Web Site

Using Open Calais you can provide search engine crawlers with rich semantic data to consider when they index your page. The Yahoo! search engine analyzes semantic data provided in Microformats, and other search engines are likely to follow.

The MOSS 2007 Open Calais Integration allows you to attach the Microformats data to dynamic content pages on your site with no development effort on your

part. As a result, users accessing your web site through search engines will get better-targeted results.

2. How It Works

When a content page is written in the web site, the author is using versioning commands such as check-in and publish to save the page. The MOSS 2007 Open Calais Integration catches these events and invokes the Open Calais Web Service to retrieve rich semantic metadata for the page. This data is saved alongside the content page.

In order to enhance search engines, the MOSS 2007 Open Calais Integration identifies page requests made by search robots. For requests made by browsers, your content page will be returned unchanged.

When a search robot is identified, the MOSS 2007 Open Calais Integration injects the Microformats retrieved for the requested page into the original content page and returns the result to the search robot.

Search engines that analyze the Microformats can offer intelligent search for your web site.

2.1. Example

The following is a sample content page written in MOSS 2007 publishing site.

The screenshot shows a web browser displaying a page from a MOSS 2007 publishing site. The page title is "MOSS 2007 OpenCalais Integration" under the "Press Releases" section. The content includes a news item from CNN about Sen. Barack Obama's announcement as the Democratic nominee for the 2008 presidential election. The page features a search bar, navigation links, and a footer with a YAHOO! Search Enabler by Thomson Reuters OpenCalais link.

publishing > Press Releases > MOSS 2007 OpenCalais Integration

Press Releases
MOSS 2007 OpenCalais Integration

WASHINGTON (CNN) -- In what he called a "defining moment for our nation," Sen. Barack Obama on Tuesday became the first African-American to head the ticket of a major political party.

Sen. Barack Obama on Tuesday told supporters he will be the Democratic nominee.

Obama's steady stream of superdelegate endorsements, combined with the delegates he received from Tuesday's primaries, put him past the 2,118 threshold, CNN projects.

"Tonight we mark the end of one historic journey with the beginning of another -- a journey that will bring a new and better day to America," he said.

"Tonight, I can stand before you and say that I will be the Democratic nominee for president of the United States."

Obama's rally was at the Xcel Energy Center in St. Paul, Minnesota -- the same arena which will house the 2008 Republican National Convention in September

YAHOO! Search Enabler by Thomson Reuters OpenCalais

When a browser requests the page, it will be returned as is. However, when a search robot requests the page, the following result will be sent instead.

The screenshot shows a web browser window displaying a search result. The browser's address bar shows the URL: `publishing > Press Releases > MOSS 2007 OpenCalais Integration`. The page content includes a navigation menu with 'Press Releases' and 'MOSS 2007 OpenCalais Integration'. The main content area displays a news article snippet from CNN, starting with 'WASHINGTON (CNN) -- In what he called a "defining moment for our nation," Sen. Barack Obama on Tuesday became the first African-American to head the ticket of a major political party.' Below the article, there is a search bar with the text 'YAHOO! Search Enabler by Thomson Reuters OpenCalais' and a search result snippet: 'CNNYAHOO!Thomson ReutersBarack Obama'. The search result snippet is circled in red.

The semantic metadata is injected to the page.

3. Using MOSS 2007 Open Calais Integration

3.1. License

The Calais MOSS 2007 Open Calais Integration is open-source.

3.2. System Requirements

The MOSS 2007 Open Calais Integration web server requirements are as follows:

1. Windows Server 2003/2008
2. Microsoft Office SharePoint Server 2007 Standard/Enterprise
3. The web server should be allowed to establish outbound connections to <http://api.opencalais.com/>. Some web hosting services disable outbound connections for security reasons, and allow such connections only upon specific request.

3.3. Installation

Run the setup.exe or Open CalaisSetup.msi file directly and follow the on-screen directions.

The wizard will install a new SharePoint solution to the farm and will prompt you to deploy the solution to all site collections. If you choose not to deploy the solution, you will have to do so manually for each site collection from the Solution Management screen in Operations page of the Central Administration.

3.4. Configuration

When setup is complete, the Settings screen will appear. This screen can also be started from the Programs menu, under Thomson Reuters.

Thomson Reuters OpenCalais Settings

General Preferences

- Allow distribution of RDF information
- Allow generated RDF information to participate in search
- Send the URL of the document as base URL

Relative Tag Base URL

Web Service URL

Robots

Edit the following list of known crawl robots to send microformats to them along with the contents of the page. (One per line)

API Key

RDF Information is generated through a Web Service. An API Key is required to access the OpenCalais Web service. If you do not have an API key, please log in to <http://www.opencalais.com/apikey> and register for a valid key.

OK Cancel

API Key – Enter your Open Calais API key in the text box. To obtain an API key please see <http://opencalais.com/APIKey>. It is recommended that you leave other values at their original values. If you wish to make changes, the following list explains some of the values.

- Relative Tag Base URL – This prefix will be attached to rel-tags returned by the 2007 Open Calais Integration. See the following section.
- Web service URL – This URL is the address of the web service in case you want to work with another web service ("beta" for example) this is the place to change the configuration.
- Robots – Use this list to tell MOSS 2007 Open Calais Integration how to identify search crawl robots. Each robot sends a specific value in the HTTP USER-AGENT Header. When the MOSS 2007 Open Calais Integration detects a robot, the Microformats data is inserted in the page.

3.5. Relative Tag Base URL

The Relative Tag Base URL is used as the base URL for Open Calais generated Rel-Tags microformats. (More information about Open Calais Microformats support can be found in the [Open Calais API Documentation](#).)

Rel-tags generated by Open Calais meet the rel-tag definition (<http://microformats.org/wiki/rel-tag>) and appear in this format:

```
<a href="http://www.YourWebSite.com/YourPath/TagName"
rel="tag">Tag</a>
```

Note that this URL **must** exist, therefore the part provided by you (www.YourWebSite.com/YourPath/), which is the value of Relative Tag Base URL, along with the Open Calais generated tag must create a valid URL.

In many cases this link would look like this:

```
http://www.MyWebSite.com/Tags?name=CalaisGenerateTag
```

Where Relative Tag Base URL is set to www.MyWebSite.com/Tags?name= Calais adds/attaches the suggested tag.

3.6. Activation

In order to activate MOSS 2007 Open Calais Integration for your web site, you must follow the instructions here:

1. Activate the Microsoft Office SharePoint Server Publishing Infrastructure feature. Browse to the site collection's site settings page, click the site collection features link under the site collection administration section and make sure the Office SharePoint Server Publishing Infrastructure feature is activated.
2. Activate the Thomson Reuters Open Calais Semantic Web Add-In feature. browse to the site settings page of your web site, click the site features link under the site administration section and activate the Thomson Reuters Open Calais Semantic Web Add-In feature.
3. Add the Yahoo! Search Enabler by Thomson Reuters Open Calais Web-part to Content pages. Create a new page by selecting Create Page from the Site Actions menu, choose Page Layout that supports web parts. Add the web-part by clicking Add a Web Part in one of the web part zones. In the web part list select the YAHOO! Search Enabler by Thomson Reuters Open Calais Web-part and select Add.

You may want to hide the text "YAHOO! Search Enabler by Thomson Reuters Open Calais Web-part" from appearing in your page, you can do

so by selecting "Modify Shared Web-Part" from the web-part's Edit menu and selecting "None" in the Chrome Type under the Appearance section.

3.7. Verification

Verify your installation and activation by performing the following steps:

1. Browse to a page to which you added the web-part; you should see the page as you did before.
2. Now browse to the same page and add rdf parameter with the value 1 (e.g., enter <http://your-domain/pages/somepage.aspx?rdf=1>). You should see the page with Microformats data injected where the web-part is located.

3.8. Troubleshooting

In the event of an error, the MOSS 2007 Open Calais Integration writes errors to the server's application event log. You can identify those events by their source (Thomson Reuters Open Calais Web Add-In).

If no events are written, make sure that the user running the Application Pool of the web application has sufficient privileges to access the Event Viewer.

If the problems occur in MOSS 2007, more information can be found in the MOSS log in the "%ProgramFiles%\Common Files\Microsoft Shared\web server extensions\12\LOGS" folder.