



Memoori News Search XML Specification Document

Document Version: 1.2

Date: March 2009

Author: Jim McHale (jim@memoori.com)

Contents

Overview	3
Memoori News Search Requests	4
Query Terms.....	4
Request Parameters.....	4
URL Escaping.....	4
Memoori XML Results DTD	6
Escaped Characters	6
Example XML Result for a Sample Search Query.....	7
Detailed List of Search Results XML Tags.....	8
Resources.....	11
XML / XSLT Translation	11
Web Development Frameworks & Cross-Browser JavaScript Libraries	11

Overview

The Memoori News Search service enables customers to display industry news articles on their own web sites. The News Search service uses a simple interface to serve news results over HTTP. Administrators have control over the way they request news results and the way they present those results to the end user. This document describes the technical details of the Memoori news requests and results formats.

To retrieve Memoori News Search results, your application must send Memoori a simple HTTP request. Memoori then returns news results in XML format. XML-formatted results give you the ability to customize the way search results are displayed.

The XML search results that Memoori returns can be displayed as HTML on your Web pages by using XSLT (<http://en.wikipedia.org/wiki/XSLT>) to transform XML into HTML. There are numerous other ways to display our results. Implementations can be found in nearly all programming languages and Web frameworks. See the 'Resources' section of this document for links to popular examples.

Memoori News Search Requests

The Memoori news search request is a standard HTTP GET command. It includes a distinct URL (which includes your unique client ID) and the parameters relevant to your query. These parameters are included in the request URL as “name=value” pairs separated by ampersand (&) characters. Parameters include data like the search query. The News Search service returns XML results in response to your HTTP requests.

Query Terms

Search query parameter values must be URL-escaped. Note that you will need to substitute the plus sign (“+”) for any whitespace sequences in the search query. This is discussed further in the URL Escaping section of this document. The search query term is submitted to the News Search service using the “q” parameter. A sample search query term would be:

```
q=oakdene+housebuilder
```

Request Parameters

This section lists the parameters that you can use when making a search request.

- **q** – the search term.
- **num** – the requested number of search results.
- **start** – the number of the search result you want to start from.

The example below shows a simple Web Search HTTP request where 7116e119732fc3e4ccc14555321bg78g is your unique client ID;

```
http://75.112.255.10/VolgaWS/services/base/news/7116e119732fc3e4ccc14555321bg78g/  
q=oakdene+housebuilder&start=0&num=10
```

The request asks for the first 10 results (start=0&num=10) for the query term "oakdene housebuilder".

By default, just requesting the q parameter (q=oakdene+housebuilder) will return the first 10 results.

URL Escaping

HTTP URL requests can only contain certain characters and therefore be aware that you should URL escape all non-alphanumeric characters in your search requests.

To URL escape a string convert all whitespace characters to a single “+” character e.g. “oakdene housebuilder” becomes “oakdene+housebuilder”.

The table below shows all the characters on the left that should be URL escaped and displayed as the encoding on the right.

Character	Hexadecimal Encoding
\$	%24
-	%2D
_	%5F
.	%2E
!	%21
*	%2A
"	%22
'	%27
(%28
)	%29
;	%3B
/	%2F
?	%3F
:	%3A
@	%40
	%7C

Memoori XML Results DTD

Memoori uses the same DTD file to describe the XML format for all its search results. Some tags in the XML only apply to certain search queries. As a result, the definitions in the DTD may be less restrictive than the definitions given in this document.

This document describes those aspects of the DTD that are **currently** relevant for News Search. When you look at the DTD, please use this document as a reference guide. If the definition differs between the DTD and the documentation, that fact is noted in this document. Some tags or elements in the DTD may not be described in this document because they are not yet available and will form part of future upgrades to the service.

The DTD is a guide to help administrators and XML parsers understand Memoori's XML results. Because Memoori's XML grammar may change from time to time, you should not configure your parser to use the DTD to validate each XML result.

Additionally, you should not configure your XML parser to fetch the DTD each time you submit a search request. Memoori updates the DTD very infrequently, and these requests will only create unnecessary delay.

You can access the latest DTD at <http://www.memoori.com/memoori.dtd>.

Escaped Characters

Certain characters must be escaped when included as values in XML tags. Your XML processor should convert these entities back to the appropriate characters. If you do not convert entities properly, the browser may, for example, render the & character as "&". The XML Standard documents these characters here (<http://www.w3.org/TR/1998/REC-xml-19980210#dt-escape>), we have included a table below for ease of reference:

	Character	Entity	Character Code
Ampersand	&	&	&
Single Quote	'	'	'
Double Quote	"	"	"
Greater Than	>	>	>
Less Than	<	<	<

Example XML Result for a Sample Search Query

Our example is taken from the News Search Query below;

```
http://75.112.255.10/VolgaWS/services/base/news/7116e119732fc3e4ccc14555321bg78g/
q=oakdene+housebuilder
```

The request asks for the first 10 results from the search term "oakdene housebuilder".

This request produces the XML result below. We have included several comments in the XML to make it easier for you to understand;

```
<?xml version="1.0" encoding="UTF-8"?>

<MSP VER="1.0">
<TM>0.036</TM>
<Q>oakdene housebuilder</Q>
<PARAM name="q" value="oakdene housebuilder" original_value="oakdene+housebuilder"/>
<PARAM name="num" value="10" original_value=""/>
<PARAM name="start" value="0" original_value=""/>

<RES>
<M>69</M>
<NB>
<PU/>
<NU>/7116e119732fc3e4ccc14555321bg78g/q=oakdene+housebuilder&start=10</NU>
</NB>

<R N="1">
<U>http://www.propertyweek.com/story.asp?sectioncode=297&storycode=3132172&c=1</U>
<T>Housebuilder Oakdene faces collapse</T>
<category>Residential news from Property Week</category>
<S>Oakdene Homes, the listed housebuilder, is on the brink of collapse.</S>
<text>Skip:[ To Main Navigation | Secondary Navigation | Third Level Navigation | Page Content |
Site Search ] Search Property Week keyword Advanced Search 29.01.09 PW Home News By region
By sector Today's papers Global Events 1st Friday Awards Conferences Events Calendar Global
Investment Club MAPIC 2008 MIPIM 2008 Offices 08 Pic of the Week RESI 08 Submit an event The
Morning After Video & Audio Deals+Data Deals Search Experian Town Reports Market Reports
Property Investment Data People People News 1st Friday Your comments Soapbox Finance Latest
News City View Derivative Values Below the Radar Your Property Latest News APC APC Advice
Sustainable News Energy Performance - EPC Sustainability TV Case Studies Blog Jobs Skip to [ Story
Content and jump story attachments ] Newsletter Sign-up Breaking News Alert Daily News Global
News People Newsletter Weekly Newsletter - - - - - </text>
<lastmodified>22 Jan 2009 14:28:00 GMT</lastmodified>
</R>
/* The results in this example would include 9 more results, each enclosed by an R tag. */
</RES>
</MSP>
```

Detailed List of Search Results XML Tags

A complete list of the XML tags currently used by the Memoori News Search service are explained in the tables below;

? means the XML tag is optional

*means there may be 0 or more instances of the XML tag

+ means there may be 1 or more instances of the XML tag

MSP	
Definition	The <MSP> tag encapsulates all data returned in the Memoori XML news results. MSP is an abbreviation of "Memoori Search Protocol".
Attributes	VER – This attribute specifies the version of the search results output. The current version is "1.0".
Example	<MSP VER="1.0">
HAS SubTags	PARAM+, Q, RES?, TM
Content Format	Empty

TM	
Definition	The <TM> tag identifies the time needed to return the search results. It is measured in seconds.
Example	<TM>0.036</TM>
SubTag OF	MSP
Content Format	Text – Floating Point Number

Q	
Definition	The <Q> tag identifies the search query submitted by the user in their HTTP request and is therefore the basis of the XML results.
Example	<Q> oakdene housebuilder </Q>
SubTag OF	MSP
Content Format	Text

PARAM	
Definition	The <PARAM> tag identifies the input parameters submitted as part of the users HTTP request. There is one PARAM tag for each parameter submitted.
Attributes	NAME – the name of the input parameter VALUE – the actual value of the parameter ORIGINAL_VALUE – original value of the input parameter (URL Escaped)
Example	<PARAM NAME="start" VALUE="0" ORIGINAL_VALUE=""/>
SubTag OF	MSP
Content Format	Complex

RES	
Definition	The <RES> tag contains the set of news results and meta-data associated with those results.

Example	<RES>
SubTag OF	MSP
Has SubTags	M, NB?, R*
Content Format	Empty

M	
Definition	The <M> tag identifies the estimated total number of results returned by the users HTTP query.
Example	<M>69</M>
SubTag OF	RES
Content Format	Text

NB	
Definition	The <NB> tag contains navigation data – i.e. the links to the next page or previous page of news result sets.
Example	<NB>
SubTag OF	RES
HAS SubTags	NU?, PU?
Content Format	Empty

PU	
Definition	The <PU> tag identifies the URL link to the previous page of news results, if such a page exists.
Example	<PU> /7116e119732fc3e4ccc14555321bg78g/q=oakdene+housebuilder&start=10</PU>
SubTag OF	NB
Content Format	Text (URL)

NU	
Definition	The <NU> tag identifies the URL link to the next page of news results, if such a page exists.
Example	<NU> /7116e119732fc3e4ccc14555321bg78g/q=oakdene+housebuilder&start=10</NU>
SubTag OF	NB
Content Format	Text (URL)

R	
Definition	The <R> tag contains the data for an individual news result.
Attributes	N – indicates the index number of the search result
Example	<R N="1">
SubTag OF	RES
Has SubTags	U, T, S, text, indexdate, lastmodified, category
Content Format	Empty

U	
Definition	The <U> tag contains the URL of the news article.
Example	<U> http://www.propertyweek.com/story.asp?sectioncode=297&storycode=3132172&c=1 </U>
SubTag OF	R
Content Format	Text (URL)

T	
Definition	The <T> tag contains the title of the news article.
Example	<T> Housebuilder Oakdene faces collapse</T>
SubTag OF	R
Content Format	Text (HTML)

S	
Definition	The <S> tag contains a summary of the document.
Example	<S> Oakdene Homes, the listed housebuilder, is on the brink of collapse.</S>
SubTag OF	R
Content Format	Text (HTML)

text	
Definition	The <text> tag contains text extracted from the news article HTML at the URL given in the <U> tag.
Example	<text>Skip:[To Main Navigation Secondary Navigation Third Level Navigation Page Content...</text>
SubTag OF	R
Content Format	Text (HTML)

lastmodified	
Definition	The <lastmodified> tag gives the date and time in the GMT time zone when the author of the news article last modified it.
Example	<lastmodified>22 Jan 2009 14:28:00 GMT</lastmodified>
SubTag OF	R
Content Format	Date Time

category	
Definition	The <category> tag contains a description of the source of the news article.
Example	<category>Residential news from Property Week</category>
SubTag OF	R
Content Format	Text (HTML)

Resources

XML / XSLT Translation

XML Wikipedia Article

<http://en.wikipedia.org/wiki/XML>

Extensible Markup Language (XML) 1.0

<http://www.w3.org/TR/1998/REC-xml-19980210>

DTD Wikipedia Article

http://en.wikipedia.org/wiki/Document_Type_Definition

XML Quick Reference Guide

<http://www.mulberrytech.com/quickref/XMLquickref.pdf>

XSL Transformations Wikipedia Article

<http://en.wikipedia.org/wiki/XSLT>

XML Template Engine Wikipedia Article

http://en.wikipedia.org/wiki/XML_template_engine

Xalan-Java – Open Source XSLT processor for transforming XML documents into HTML.

<http://xml.apache.org/xalan-j/>

Google AJAXSLT is an implementation of XSLT in JavaScript, intended for use in AJAX applications.

<http://code.google.com/p/ajaxslt/>

Web Development Frameworks & Cross-Browser JavaScript Libraries

Django – A Python Open Source Web Framework.

<http://www.djangoproject.com/>

Ruby on Rails – Open Source Web Framework.

<http://rubyonrails.org/>

Microsoft Silverlight - A programmable web browser plugin.

<http://www.microsoft.com/silverlight/>

Adobe Flex – Cross Browser Platform for development based on the proprietary Adobe Flash.

<http://flex.org/>

Prototype – A JavaScript Framework for development of Web applications.

<http://www.prototypejs.org/>

Extjs – A cross-browser JavaScript library

<http://extjs.com/>