

Mergeflow.net Reader API Specification

Contact	Markus Schmidt markus.schmidt@microstep-it.de , +49 89 4377779-124
API Version	0.3
Specification Revision	1
Status	Draft

History

Date	Author	Rev.	Remarks
2011-04-19	Markus Schmidt	1	Added User resource model as a suggestions for the use case "enumerating a user's profiles"
2011-04-20	Tobias Eichner	1	Added new Profile attributes and elements.

Content

Abstract.....	4
Examples for Use Cases	4
Current Limitations	4
Common Behaviors	4
Transport Protocol	4
URI Space	4
Media Types	4
Request Headers.....	4
Request Parameters	5
Response Headers	5
Standard HTTP Status Codes	5
Error Response Message Bodies.....	5
Resource Models	6
Profile	6
Keyword	8
Document.....	8

Abstract

Mergeflow.net users can create topic profiles based on web feeds. The mergeflow.net website offers different views on these topic profiles, e. g. different kinds of grouping, sorting, clustering, and tag clouds.

This document specifies an API for accessing the underlying data of these views in a structured way.

Examples for Use Cases

- Integration of topic profile data into desktop applications
- Integration of topic profile data into web or intranet sites, using a custom layout and design
- Smartphone applications optimized for mobile usage

Current Limitations

The underlying data of the following topic profile views are not available: Past, Feeds, and Sites.

For every request, the number of items returned is limited. Currently there is no way to retrieve data beyond those limits.

At this time, user authentication is not covered, so the API cannot be used to access data of private (non-public) topic profiles.

This API cannot be used to create or modify topic profiles.

Common Behaviors

The following constraints apply to all requests and responses that occur in this RESTful API.

Transport Protocol

This API is based on the Hypertext Transfer Protocol, version 1.1 (RFC 2616).

URI Space

The resources are Profiles, Keywords, and Documents. All are identified by URIs. To begin operations, a client must know the URI for a profile. The profile URI can be extracted from the HTML representation of a profile view: It will contain an HTML link element with media type *application/vnd.mergeflow.net.ReaderAPI.Profile+xml*.

Example:

```
<link rel="ReaderAPI" type="application/vnd.mergeflow.net.ReaderAPI+xml" href="http://mergeflow.net/readerapi/profiles/ms/itsec" />
```

Dereferencing the profile URI will yield a representation of the profile containing links to the recent keywords and documents in that profile.

Clients must not make assumptions about the layout of the URIs within a topic profile.

Media Types

In the Platform API, resource representations are encoded in XML.

Each type of resource has its own media-type, which matches the pattern *application/vnd.mergeflow.net.ReaderAPI.XXX+xml*, where "XXX" represents the portion of the identifier unique to a particular representation format for each service.

Request Headers

Beside the basic required HTTP headers like *Content-Length* and *Host* several specific HTTP headers are used as described in the following table:

Name	Supported Values	Description	Required
Accept	Comma-delimited list of media types	Indicates to the server what media type(s) this client is prepared to accept.	Recommended. If the list contains the according Reader API media type (see "Media Types" above), the content type of the response will be set to the correct content type. Otherwise, the content type will be <i>application/xml</i> . This behavior is intended to support web browsers as user agents for testing and evaluation.
X-MergeflowNet-ReaderAPI-Specification-Version	String containing a specification version number.	Declares the specification version of the API that this client was programmed against.	No (current version is assumed if this header is not present).

Request Parameters

Since the URI space is controlled by the server, client programs must not make any assumptions about the meanings of request parameters.

Response Headers

Beside the basic required HTTP headers like *Content-Length* several specific HTTP headers are used as described in the following table:

Name	Supported Values	Description
Content-Type	Media type describing the response message body.	Describes the representation and syntax of the response message body. If the user agent included the proper Reader API media type (see "Media Types" above) in the <i>Accept</i> request header, the specific media type will be returned, otherwise <i>application/xml</i> to support web browsers as user agents for testing and evaluation.
X-MergeflowNet-ReaderAPI-Specification-Version	String containing a specification version number.	This response was generated according to the specified version of the API specification.
X-MergeflowNet-ReaderAPI-Implementation-Version	String containing a specification version number.	Version number of the application running the API. This value could be useful when evaluating errors and problems with Mergeflow's technical support.

Standard HTTP Status Codes

The API will return standard HTTP response codes as described in the following table:

HTTP Status	Description
200 OK	The request was successfully completed.
400 Bad Request	The request could not be processed because it contains missing or invalid information, or the specification version is not supported. The response message body will contain more detailed information.
404 Not Found	The request specified a URI of a resource that does not exist or is not publicly available.
500 Internal Server Error	The server encountered an unexpected condition which prevented it from fulfilling the request.
503 Service Unavailable	The server is currently unable to handle the request due to temporary overloading or maintenance of the server.

Error Response Message Bodies

If a response is returned with an error status code (400-499 or 500-599), the server returns a response message body containing an Error data model, containing the error code and a message describing what went wrong. The text value of such messages might be used, for example, to communicate with a human user of the client side application.

The media type is *application/vnd.mergeflow.net.ReaderAPI.Error+xml*.

Resource Models

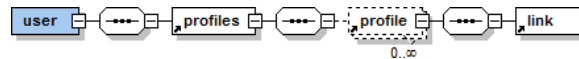
This section specifies the representations of the resources on which this API operates. Attached to this specification, an XML Schema is available for each resource.

Each type of resource has its own Internet Media Type. The media type conforms to the pattern *application/vnd.mergeflow.net.ReaderAPI.XXX+xml*.

User

Media Type: *application/vnd.mergeflow.net.ReaderAPI.User+xml*

A user resource represents a user's list of public profiles.



The *profile* root element contains the following attributes and elements:

Name	Type	Occurs	Description
alias	xs:string	1	The user's alias as it occurs in the URL.
profiles		1	Parent element for <i>profile</i> elements, listing the currently active public profiles of the user.
profiles/profile		1..n	Describes a currently active and public topic profile.

Note: Requesting a user that has not published any profile, the result will be the same as when requesting a non-existing user, HTTP 404 (Not Found). This is for security reasons. Therefore, if a user resource is returned, there will be at least one public profile listed.

The *profile* element contains the following attributes and elements:

Name	Type	Occurs	Description
title	xs:string	1	The topic profile title as specified by the profile owner.
description	xs:string	1	The topic profile description as specified by the profile owner.

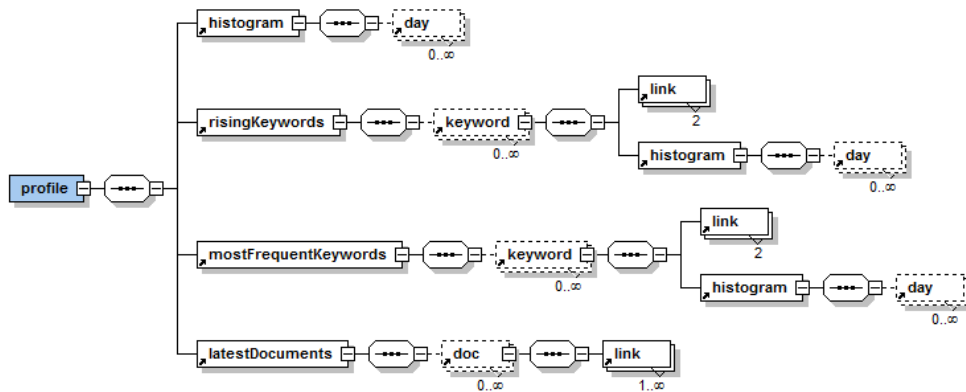
The *link* element contains the following attributes:

Name	Type	Occurs	Description
rel	xs:NMTOKEN	1	The link element with a rel attribute value of "recent" is referencing the profile URI of a profile, providing access to a profile's master data, current keywords and latest documents.
href	xs:anyURI	1	The URI of the resource.
type	xs:string	1	The media type of the resource.

Profile

Media Type: *application/vnd.mergeflow.net.ReaderAPI.Profile+xml*

A profile resource represents a profile's master data, keywords and recent documents.



The *profile* root element contains the following attributes and elements:

Name	Type	Occurs	Description
title	xs:string	1	The topic profile title as specified by the profile owner.
description	xs:string	1	The topic profile description as specified by the profile owner.
snapshotdaycount	xs:int	1	The number of days covered in the current view on this topic profile.
documentsToday	xs:int	1	The number of documents found today (00:00 UTC until 'now').
histogram		1	Parent element for <i>day</i> elements, containing the number of documents per day.
risingKeywords		1	Parent element for <i>keyword</i> elements, listing the currently rising keywords found in the topic profile.
risingKeywords/keyword		0..n	Describes a currently rising topic profile keyword.
mostFrequentKeywords		1	Parent element for <i>keyword</i> elements, listing the 50 most frequent keywords found in the topic profile.
mostFrequentKeywords/keyword		0..n	Describes a most frequent profile keyword.
latestDocuments		0..n	Parent element for <i>doc</i> elements, listing the latest documents of the profile (unfiltered).
latestDocuments /doc		0..n	Describes a latest document.

The *keyword* element contains the following attributes and elements:

Name	Type	Occurs	Description
name	xs:string	1	The keyword itself.
weight	xs:double	1	Current weight of the keyword, based on the number of matching documents. Value is in the range]0..1].
histogram		1	Parent element for <i>day</i> elements, containing the number of documents matching this keyword per day. The number of <i>day</i> elements matches the <i>snapshotdaycount</i> attribute of the <i>profile</i> element.
link		2	Link elements referencing URIs yielding a Keyword resource, listing matching documents for the keyword. Two variants of this resource are available: One with a <i>rel</i> attribute of "latestDocs" specifying the URI yielding the list of the most recent documents matching the keyword, and one with a <i>rel</i> attribute of "relevantDocs" specifying the URI yielding the list of the most relevant documents for the keyword.

The *day* element contains the following attributes:

Name	Type	Occurs	Description
offset	xs:int	1	Number of days counting backwards from today.
docCount	xs:int	1	The number of documents related to this keyword on that day.

The *link* element contains the following attributes:

Name	Type	Occurs	Description
rel	xs:NMTOKEN	1	One of the following values: details, latestDocs, original, relevantDocs. See the according table for an explanation of these values in the context of the link's parent element.
href	xs:anyURI	1	The URI of the resource.
type	xs:string	1	The media type of the resource.

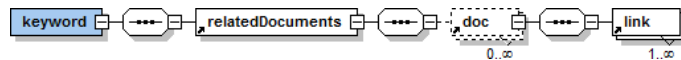
The *doc* element contains the following attributes and elements:

Name	Type	Occurs	Description
title	xs:string	1	Original document title
teaser	xs:string	1	Short excerpt of the original document, can be used as a preview text.
downloadTimestamp	xs:dateTime	1	Date and time when Mergeflow retrieved this document.
firstReportedTimestamp	xs:dateTime	1	Date and time when the first web feed item in this topic profile reported the document's URI.
link		2	The link element with a rel attribute value of "original" is referencing the original URI of the document. The link element with a rel attribute value of "details" is referencing an URI yielding a Document resource, listing document details as well as a list of similar documents.

Keyword

Media Type: *application/vnd.mergeflow.net.ReaderAPI.Keyword+xml*

A keyword resource represents a rising keyword of the profile and contains a list of related documents (according to the keyword).



The *keyword* root element contains the following attributes and elements:

Name	Type	Occurs	Description
name	xs:string	1	The keyword itself.
relatedDocuments		1	Parent element for <i>doc</i> elements, listing the documents tagged with the keyword.
relatedDocuments/doc		0..n	Describes a related document.

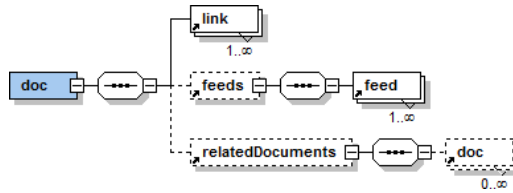
The *doc* root element contains the following attributes and elements:

Name	Type	Occurs	Description
relevance	xs:double	1	The relevance of the document according to the keyword. Value is in the range]0..1].
title	xs:string	1	Original document title.
teaser	xs:string	1	Short excerpt of the original document, can be used as a preview text.
downloadTimestamp	xs:dateTime	1	Date and time when Mergeflow retrieved this document.
firstReportedTimestamp	xs:dateTime	1	Date and time when the first web feed item in this topic profile reported the document's URI.
link		1..n	The link element with a rel attribute value of "original" is referencing the original URI of the document. The link element with a rel attribute value of "details" is referencing an URI yielding a Document resource, listing document details as well as a list of similar documents.

Document

Media Type: *application/vnd.mergeflow.net.ReaderAPI.Document+xml*

A document resource represents the data of the original document downloaded by Mergeflow and (if available) a list of feeds which reported the document.



The `doc` root element contains the following attributes and elements:

Name	Type	Occurs	Description
title	xs:string	1	Original document title.
teaser	xs:string	1	Short excerpt of the original document, can be used as a preview text.
downloadTimestamp	xs:dateTime	1	Date and time when Mergeflow retrieved this document.
relevance	xs:double	0..1	The relevance of the document according to a keyword or a document. Value is in the range [0..1].
firstReportedTimestamp	xs:dateTime	0..1	Date and time when the first web feed item in this topic profile reported the document's URI.
link		1..n	The link element with a rel attribute value of "original" is referencing the original URI of the document. The link element with a rel attribute value of "details" is referencing an URI yielding a Document resource, listing document details as well as a list of similar documents.
feeds		0..n	Parent element for <code>feed</code> elements, listing the feeds which reported the document.
feeds/feed		1..n	Describes a feed which reported the document.
relatedDocuments		1	Parent element for <code>doc</code> elements, listing the documents similar to the current one.
relatedDocuments/doc		0..n	Describes a related document.

The `feed` root element contains the following attributes:

Name	Type	Occurs	Description
name	xs:string	1	The name of the feed.
reportedTimestamp	xs:dateTime	1	Date and time when the feed reported the document.