## Wikipedia API Functionality

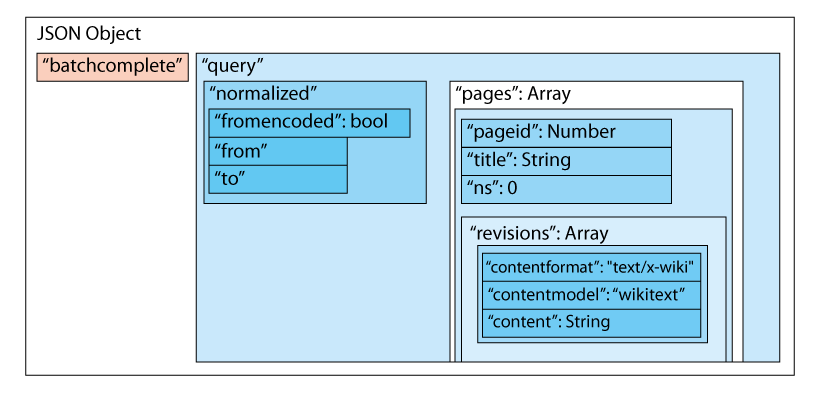
(Note that [RestBase](https://www.mediawiki.org/wiki/RESTBase) or [Wikidata query](https://www.mediawiki.org/wiki/Wikidata_query_service) API could be used instead – with certain features different)

The wikipedia API endpoint url: <https://en.wikipedia.org/w/api.php>. All request urls sent to Wikipedia API should contain it, with some additional parameters, which define your request.

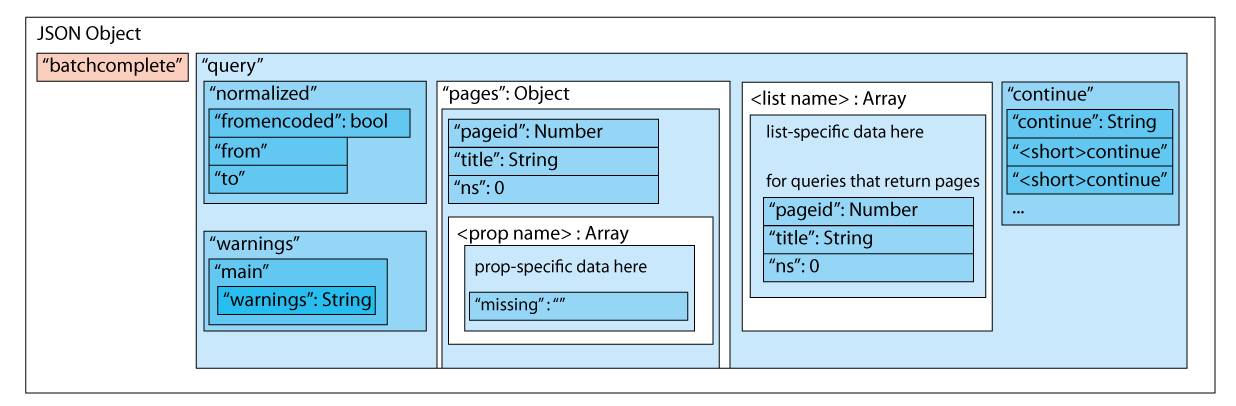
|  |  |  |
| --- | --- | --- |
| Parameter | Possible values | Description |
| action | query || edit || help | What action do you want to perform.  (edit requires logging in to Wikipedia as a user)  In this project only query will be used – to get information.  More can be found on: [Media Wiki API help](https://en.wikipedia.org/w/api.php) |
| format | json || xml || jsonfm | Format type (Most of the page is still sent as text – inside json or xml)  jsonfm will return an html version of json – for debugging only |
| formatversion | 1 || 2 | 2 supports all Unicode characters. 1 encodes them other than ASCII characters in hex. |
| prop | Description: Prop is the main parameter for querying information with large amounts of text. This parameter defines exactly what query do you want to perform – get a page, its language links, etc.  Required parameters: titles, pageids or revids | |
| revisions | Gives information about page revisions. (short form - rv) |
| categories | Categories of specified pages (short form - cl) |
| contributors | Contributors of specified pages (short form - pc) |
| langlinks | Links to the page in other languates (short form – ll) |
| list | Description: List is another parameter to make your query more specific. It is used when you want to get little information about a large amount of pages. | |
| backlinks | All the pages that link to current (short form - bl) |
| allpages | List of all the pages in Wikipedia (language specific) (short form - ap) |
| allrevisions | List all the revisions in Wikipedia (refers to all pages in Wikipedia. To get revisions for a specific page use prop=revisions) (short form - arv) |
| titles | page name | page name is represented as the first parameter in the inner-Wikipedia linking format.  Page name should be sent in the format it is displayed in Wikipedia page and split by ‘ ‘. Because spaces are not supported by the URI formatting, they should be replaced with ‘%20’ in link.  If underscores are added instead of spaces, Wikipedia automatically replaces them and adds a “normalized” attribute to the json object returned.  Multiple pages can be received by splitting their names with a ‘|’  Example query=Microsoft|Samsung|Polar%20bear|United\_Nations |
| pageids ||  revids | page id || rev id | Works the same as titles, but user specifies the ids of pages to get or the ids of revisions made to page |
| All the parameters below are very specific in their usage and should refer to a previously used prop of list parameter. Their names should be preceded with the short form of their parent parameter (<short form><param>). For example, limit parameter for revisions would be set like this: rvlimit = 2. (Where rv = short form for revisions) | | |
| <short>prop | content || user || timestamp || ids | What information do we need to get (contents of the page, the last user who modified it…)  (Is used with prop – allowed values may differ from prop value) |
| <short>dir | newer || older | In which direction should all the information be sorted (by timestamps)  ! newer means it will be sorted from older to newer – older first |
| <short>limit | number || max | Specifies how many results do you want to get for your query |
| <short>continue |  | The value should be the same as specified in the result of the previous query (used to get more pages than the Wikipedia limit) |
| More properties can be found in these sources:  [API: Properties](https://www.mediawiki.org/wiki/API:Properties)  [API: Lists](https://www.mediawiki.org/wiki/API:Lists) | | |

**Returned JSON Objects**

This example shows a query request with parameters prop=revisions & rvprop=content:



This is a generalized diagram of all query requests to the Wikipedia API:



All the API:Property queries will have all the information inside the “pages” Object, because all of them refer to a specific page. The object representing a page will have an attribute, named exactly the same as our prop parameter value (in the revisions example, we can see that it is called ‘revisions’).

Most of the inner API:Property objects have attributes, that are named the same as <short>prop parameter in the request (For example, revisions have the ‘content’ attribute). But they may also have other attributes, which make the result more specific and may carry crucial for our request information.

All the API:List queries will be added to the “query” object as a separate attribute, because they independent from any pages may contain pages themselves. The attributes of the property and list objects may vary depending on request, but lists mostly have the default page structure inside them (with ‘pageid’, ‘title’ and ‘ns’)

Attributes explanation and possible values:

|  |  |
| --- | --- |
| batchcomplete | true or false depending on whether request completed successfully. False when the url is formatted wrongly or misses required attributes (like format).  If the title attribute is wrong (page could not be found) returns true |
| normalized | If the title is formatted wrongly or contains extra information, Wikipedia automatically fixes it and displays the fixes in the normalized Object |
| warnings  error | May contain any possible warnings. Most of the warnings are because of parameters were specified wrong. (In worst case may appear the error attribute) |
| continue | Contains a piece of text, that could be used to continue the query if the limit was reached, but there is still information on Wikipedia for the given request.  (The continue should be added as a separate parameter with its value copied from request) |
| pages | Object, containing all the pages with pageid as attribute and page Object as value.  If page wasn’t found a negative value is assigned (starting from -1, -2 and so on…) |

Note that there are pages that are only used to redirect to another page and do not have any content themselves.

They mostly look like this: “content”: “#REDIRECT [[<page name>]]”