Documentation of method *zemanta.suggest*

Suggest method allows developers to query Zemanta for contextual metadata about a given text. There are currently four main components of response: **articles**, **keywords**, **images**, **in-text links** and optional component **categories**.

Application submits text, either as HTML or plain text, and receives a number of suggestions back. Your text can have HTML markup such as bold, headings etc, but **do not send full HTML pages** since navigational elements will be considered part of content. You should only send the "pure content".

About Zemanta

Zemanta API is state of the art in text analysis.

For any input text you get entities, related images, articles, hyperlinks, and tags to use to your liking.

Start building

It's easy. Just three steps:
1. Create developer account
2. Obtain API key
3. Get sample code

This document is primarily concerned with the formating of input and out parameters of a request, for precise explanation how each suggestion is made, please refer to Zemanta API Companion.

Function input parameters

Parameter	Description	Required	Possible values
method	Method on the server	Yes	"zemanta.suggest"
api_key	Your API key	Yes	string
text	Input text (clear text or HTML)	Yes	string
format	requested output format	Yes	"xml", "json", "wnjson" or "rdfxml"
return_rdf_links	return URIs of Linking Open Data entities	No	0, 1
return_categories	categorize into specified categorization scheme	No	"dmoz" or partner ID
return_images	return related images (default is yes) *This can cause dramatic performance*	No	0, 1

	improvements		
return_keywords	return keywords (default is yes) *This can affect performance slightly positively*	No	0, 1
emphasis	terms to "emphasise", even when not present in text. All related articles are then required to have this term.	No	string
text_title **[NEW since August 2010]**	Title of the text you are sending. Helps the text understanding algorithm.	No	string
personal_scope	return only personalized related articles and images	No	0, 1
markup_limit	Number of in-text links to return (default: depending on the number of input words, 1 per each 10 words, and it maxes out at 10)	No	number
images_limit	Number of images to return (default:24)	No	number
articles_limit	Number of articles to return (default:10)	No	number
articles_max_age_days	Maximum age of returned articles (default: no limit)	No	number
articles_highlight **[NEW since August 2010]**	Should a highlighted search snippet for each article be returned, where available (default: no)	No	number
image_max_w	Maximum image width (default: 300)	No	number
image_max_h	Maximum image height (default: 300)	No	number
sourcefeed_ids	ID for personalized related articles	No	
flickr_user_id	flickr ID of the user	No	
pixie	the chosen Zemanta signature icon	No	

Parameters in green should be passed directly from the response of zemanta.preferences call. Do not try to put your own values there.

About response formats

- **xml**
  XML is the format for interchange of information on internet. Zemanta offers a simple XML response format.
- **json**
  In dynamic languages JSON is more natural format to parse, so you can use "json" as *format*, to get such response. Generally it is structured the same as "xml" format mentioned above.
- **wnjson**
  In JavaScript there are additional issues in calling Zemanta API. When sending large chunks of data you need a POST call and in order to handle that in JavaScript you need a trick called "window name json". You can open such call inside IFrame and then read the content of the window title to get to the plain JSON. Generally you should use frameworks such as jQuery that support this kind of call natively.
- **rdfxml**
  Since Zemanta is a semantic application it is expected that proper semantic response is offered. When

specifying "rdfxml" format you will get RDF/XML structure as response. We suggest using semantic libraries to read the triples encoded inside. All objects inside this response are properly typed and we have documented it on a separate page. There you can also find more information about possible uses of Zemanta in semantic software/projects/ecosystems. Response is more precisely described in an "Zemanta RDF response" document.

Response structure (top level)

Parameter	Description	Type	Possible Values
status	indicates the status of request	string	ok, fail
rid	unique request id that can be used with calls that require it	string	36 chars UUID4
articles	a list of objects	list	
keywords	a list of objects	list	
images	a list of objects	list	
markup	object	dict	
categories	a list of objects	list	optional (when using categorization)
signature	signature to use (HTML blob)	string	

Articles substructure

Articles substructure is a list of *article* objects where each object has the following format:

Article object	Description	Type
url	URL of the article	string
title	title of the article	string
published_datetime	date when article was published. If not available harvested date is used. In ISO 8601 format.	string
confidence	confidence on 0.0 to 1.0 scale	float
zemified	is the article zemified or not (1 or 0), optional	integer

Keywords substructure

Keywords substructure is a list of *keyword* objects where each object has the following format:

Keyword object	Description	Type
name	keyword (can contain spaces, but not commas)	string
confidence	confidence on 0.0 to 1.0 scale	float

schema	origin of the keyword (right now "general")	string

Images substructure

Images substructure is a list of *image* objects where each object has the following format:

Image object	Description	Type
url_l	URL of large version of the image	string
url_m	URL of medium version of the image	string
url_s	URL of small version of the image	string
url_l_w	width of large image	integer
url_l_h	height of large image	integer
url_m_w	width of medium image	integer
url_m_h	height of medium image	integer
url_s_w	width of small image	integer
url_s_h	height of small image	integer
source_url	URL of page that has more information about the image	string
license	license of image (HTML blob)	string
description	description of image (text)	string
attribution	attribution of image (HTML blob)	string
confidence	confidence on 0.0 to 1.0 scale	float

Markup substructure

Markup substructure has two substructures:

Markup object	Description	Type
text	**[this is always an empty string as of August 2010]**	string
links	a list of objects	list

Structure of each *link* object

Link object	Description	Always	Type
anchor	the word(s) in original text that should be anchored	yes	string
confidence	confidence that the disambiguation is correct	yes	float 0.0 - 1.0
relevance	relevance of the entity to the input text as a whole	yes	float 0.0 - 1.0

entity_type	a list of entity types (entity types that Zemanta knows about)	optional and can be empty	list
target	a list of objects	yes	list

Structure of each *target* object

Property	Description	Type
url	resource URL of the linked term	string
type	type of resource	string
title	title of resource	string

Type can be one of the following strings:

- amazon
- answerscom
- anyclip
- blog
- chacha
- crunchbase
- facebook
- gamerevolution
- geolocation
- googlefinance
- homepage
- hulu
- imdb
- itis (Integrated Taxonomic Information System)
- justslimming
- lastfm
- musicbrainz
- muzu
- mybloglog
- myspace
- myspaceeverything
- ncbi (National Center for Biotechnology Information)
- rdf (for semantic links to DBpedia, Freebase and MusicBrainz)
- snooth
- thecarconnection
- tracked
- twitter
- uptake
- webmd
- wikinvest
- wikipedia
- yahoofinance
- yahoosports

- youtube

Categories substructure

Categories substructure is a list of *category* objects where each object has the following format:

Category object	Description	Type
name	category name	string
confidence	confidence on 0.0 to 1.0 scale	float
categorization	what categorization this category comes from	string

If you don't have special arrangement with Zemanta you can only get "dmoz" as *categorization.*

You can also check API use sample code in different languages (PHP, Perl, C#,...).

Sample call (python)

```
import urllib

gateway = 'http://api.zemanta.com/services/rest/0.0/'
args = {'method': 'zemanta.suggest',
        'api_key': 'key1234',
        'text': '''The Phoenix Mars Lander has successfully deployed its robotic arm and
tested other instruments including a laser designed to detect dust,
clouds, and fog. The arm will be used to dig up samples of the Martian
surface which will be analyzed as a possible habitat for life.''',
        'return_categories': 'dmoz',
        'format': 'xml'}
args_enc = urllib.urlencode(args)
print urllib.urlopen(gateway, args_enc).read()
```

Sample response (Truncated for clarity)

```
<rsp>
  <status>ok</status>
  <articles>
    <article>
      <url>http://abcnews.go.com/Technology/story?id=5255072&amp;page=1</url>
      <confidence>0.048289</confidence>
```

```
    <published_datetime>2008-06-26T19:12:59Z</published_datetime>
    <title>Seeds of Life Found in Martian Soil</title>
    <zemified>0</zemified>
  </article>
  <article>
    <url>http://www.computerworld.com/action/article.do?command=viewArticleBasic&amp;articleId=9
    <confidence>0.0479</confidence>
    <published_datetime>2008-07-09T13:00:00Z</published_datetime>
    <title>NASA: Mars Lander short circuit pushes up ice test</title>
    <zemified>0</zemified>
</article>
</articles>
<markup>
  <text>The &lt;a class="zem_slink" href="http://en.wikipedia.org/wiki/Phoenix_%28spacecraft%29"&
  <links>
    <link>
      <confidence>0.084166</confidence>
      <anchor>Phoenix Mars Lander</anchor>
      <target>
        <url>http://www.youtube.com/watch?v=tR91HkTZ9VY</url>
        <type>youtube</type>
        <title>Phoenix (spacecraft)</title>
      </target>
      <target>
        <url>http://en.wikipedia.org/wiki/Phoenix_%28spacecraft%29</url>
        <type>wikipedia</type>
        <title>Phoenix (spacecraft)</title>
      </target>
    </link>
    <link>
      <confidence>0.006165</confidence>
      <anchor>robotic arm</anchor>
      <target>
        <url>http://en.wikipedia.org/wiki/Robotic_arm</url>
        <type>wikipedia</type>
        <title>Robotic arm</title>
      </target>
    </link>
  </links>
</markup>
<images>
  <image>
    <description>PASADENA, CA - MAY 25:  Phoenix principal investigator, University of Arizona, I
    <attribution>Image by &lt;a href="http://www.daylife.com/source/Getty_Images"&gt;Getty Images
    <license>Low resolution use allowed when backlinking</license>
    <source_url>http://www.daylife.com/image/097c92HarS9oN</source_url>
    <confidence>0.5</confidence>
    <url_s>http://cache.daylife.com/imageserve/097c92HarS9oN/75x75.jpg</url_s>
    <url_s_w>75</url_s_w>
```

```
      <url_s_h>75</url_s_h>
      <url_m>http://cache.daylife.com/imageserve/097c92HarS9oN/150x100.jpg</url_m>
      <url_m_h>113</url_m_h>
      <url_m_w>150</url_m_w>
      <url_l>http://cache.daylife.com/imageserve/097c92HarS9oN/150x100.jpg</url_l>
      <url_l_h>100.0</url_l_h>
      <url_l_w>150</url_l_w>
    </image>
    <image>
       <description>Day 2 14.19.40 Phoenix Mars Lander 3-D Anaglyphs</description>
       <attribution>Image by &lt;a href="http://www.flickr.com/photos/48836503@N00/2530611038"&gt;
       <license>License CreativeCommons Attribution only</license>
       <source_url>http://www.flickr.com/photos/48836503@N00/2530611038</source_url>
       <confidence>0.5</confidence>
       <url_s>http://farm4.static.flickr.com/3043/2530611038_f490407155_s.jpg</url_s>
       <url_s_w>75</url_s_w>
       <url_s_h>75</url_s_h>
       <url_m>http://farm4.static.flickr.com/3043/2530611038_f490407155_m.jpg</url_m>
       <url_m_w>220</url_m_w>
       <url_m_h>240</url_m_h>
       <url_l>http://farm4.static.flickr.com/3043/2530611038_f490407155.jpg</url_l>
       <url_l_w>458</url_l_w>
       <url_l_h>500</url_l_h>
    </image>
    <image>
      <description>An artist's rendition of the Phoenix Mars probe during landing. The sophisticate
      <attribution>Image via &lt;a href="http://commons.wikipedia.org/wiki/Image:Phoenix_landing.jp
      <license>Public domain</license>
      <source_url>http://commons.wikipedia.org/wiki/Image:Phoenix_landing.jpg</source_url>
      <confidence>0.99</confidence>
      <url_s>http://upload.wikimedia.org/wikipedia/commons/thumb/6/6a/Phoenix_landing.jpg/75px-Phoe
      <url_s_w>75</url_s_w>
      <url_s_h>69</url_s_h>
      <url_m>http://upload.wikimedia.org/wikipedia/commons/thumb/6/6a/Phoenix_landing.jpg/202px-Pho
      <url_m_w>202</url_m_w>
      <url_m_h>186</url_m_h>
      <url_l>http://upload.wikimedia.org/wikipedia/commons/6/6a/Phoenix_landing.jpg</url_l>
      <url_l_w>5200</url_l_w>
      <url_l_h>4800</url_l_h>
    </image>
  </images>
  <keywords>
    <keyword>
      <confidence>0.506297</confidence>
      <name>Mars</name>
      <scheme>general</scheme>
    </keyword>
    <keyword>
      <confidence>0.296248</confidence>
```

```
      <name>Phoenix</name>
      <scheme>general</scheme>
    </keyword>
  </keywords>
  <categories>
    <category>
      <confidence>0.231914</confidence>
      <categorization>dmoz</categorization>
      <name>Top/Science/Anomalies_and_Alternative_Science/Astronomy,_Alternative/Planetary_Anomalie
    </category><category>
      <confidence>0.195886</confidence>
      <categorization>dmoz</categorization>
      <name>Top/Science/Astronomy/Solar_System/Planets/Mars</name>
    </category>
  </categories>
  <signature>&lt;div class="zemanta-pixie"&gt;&lt;a class="zemanta-pixie-a" href="http://reblog.zem
  <rid>40b3d04b-5248-4256-a22b-c07ba38b2d9f</rid>
</rsp>
```


Fine print

The request size is limited. Only first 8kb of text is going to be processed.

There are also limits in place for number of requests per day (as specified in Terms of service) and per second. If you go over these limits, the system will return an error message "403 Developer over quota". Contact us if you need to make more calls to our system.

While confidence information is available for certain analysis, it is very seldom the case that comparing confidence values between documents is meaningful. Generally they represent relative measure of confidence for the specific type of recommendation for that specific document. Value should also not be interpreted as probability. We do our best to return meaningful confidences, but generally you should consult us about their use.

Zemanta

Follow us on Twitter.
Join us on Facebook.

Tech blog | Register for an API key | API | Terms of Use | Contact us