

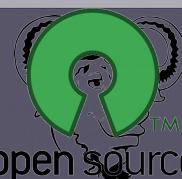
Programmiamo con Wikipedia



WIKIPEDIA
The Free Encyclopedia



Comune di Livorno





Cos'è Wikipedia ?

FabC discussioni prove preferenze beta osservati speciali contributi esci

Voce **Discussione** Leggi Modifica Modifica wikitesto Cronologia Altro ▾ Ricerca 🔎

Wikipedia

Da Wikipedia, l'enciclopedia libera.

Disambiguazione – Se stai cercando altri significati, vedi [Wikipedia \(disambigua\)](#).

Wikipedia (pronuncia: [vedi sotto](#)) è un'enciclopedia online a contenuto libero, collaborativa, multilingue e gratuito ospitata dalla Wikimedia Foundation, un lucro statunitense.

Lanciata da [Jimmy Wales](#) e [Larry Sanger](#) inizialmente nell'edizione in lingua inglese aggiunto edizioni in numerose altre lingue una [parola macedonia](#) nata dall'unione di ([da enciclopedia](#)).

Etimologicamente, Wikipedia significa "i hawaiano *wiki* (veloce), con l'aggiunta d'antico *παιδεία*, *paideia*, formazione). Wi voci in oltre 280 lingue dieci siti web più visitati consultata opera di

Ir
1 Storia
2 Descrizione
2.1 Caratteristiche

Pagina principale Ultime modifiche Una voce a caso Vetrina Aiuto Comunità Portale Comunità Bar Il Wikipediano Fai una donazione Contatti Strumenti Puntano qui Modifiche correlate Carica un file Carica su Commons Pagine speciali Link permanente Informazioni sulla pagina Elemento Wikidata Cita questa voce Stampa/esporta Crea un libro Scarica come PDF

FabC discussioni prove preferenze beta osservati speciali contributi esci

Voce Discussione Leggi Modifica Modifica wikitesto Cronologia Altro ▾ Ricerca 🔎

Wikipedia in italiano

Da Wikipedia, l'enciclopedia libera.

La **Wikipedia in italiano** è l'edizione in lingua italiana dell'enciclopedia online [Wikipedia](#). Tale edizione, che ebbe inizio ufficialmente il 11 maggio del 2001^[1], è l'ottava per numero di voci (dopo la [Wikipedia in russo](#) e prima di quella in spagnolo), con 1 307 435 pagine al 17 ottobre 2016^{[2][3]}.

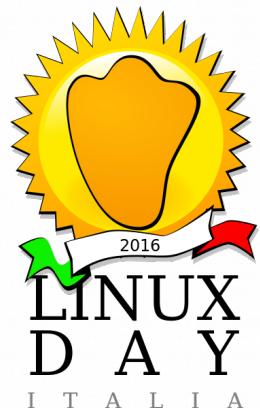
wikipedia in italiano W

wikipedia in italiano W

Statistiche [modifica | modifica wikitesto]

Al 17 ottobre 2016 l'edizione di Wikipedia in italiano conta 1 307 435 voci (con una crescita mensile di circa 10.000 unità), 4 565 167 pagine, 1 419 428 utenti registrati di cui 7 461 attivi, 112 amministratori e una "profondità" (depth) di 121.

È l'ottava Wikipedia per numero di voci ma, come "profondità", è la 15^a fra quelle con più di 100.000 voci (al 14 marzo 2016).





Contribuire a Wikipedia

- Modifica degli articoli
- Categorie di articoli
- Template: verso i dati strutturati



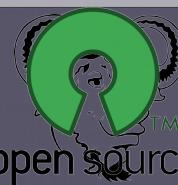


Template

```
{{Scuola
|Nome = Istituto tecnico industriale statale Galileo Galilei
|Soprannome =
|Logo =
|Dimensione logo =
|Immagine =
|Dimensione immagine =
|Nazione = ITA
|città = Livorno
|cittàlink = Livorno
|Indirizzo = Via Galilei Galileo, 66/68
|Succursali =
|Tipo = Istituto tecnico industriale
|Ordinamento = pubblico
|Fondazione = 1887
|Preside =
|Dipendenti =
```



Istituto tecnico industriale statale Galileo Galilei	
Paese	Italia
Città	Livorno
Indirizzo	Via Galilei Galileo, 66/68
<hr/>	
Tipo	Istituto tecnico industriale
Ordinamento	pubblico
Fondazione	1887
Sito	http://www.galileilivorno.gov.it/
<small>Modifica dati su Wikidata · Manuale</small>	





Accedere ai contenuti

Il motore di Wikipedia : MediaWiki



MediaWiki Action API web service

https://www.mediawiki.org/wiki/API:Main_page





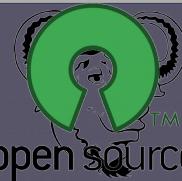
Web Service

Web service

From Wikipedia, the free encyclopedia

A **Web service** is a service offered by an electronic device to another electronic device, communicating with each other via the [World Wide Web](#). In a Web service, Web technology such as [HTTP](#), originally designed for human-to-machine communication, is utilized for machine-to-machine communication, more specifically for transferring machine readable file formats such as [XML](#) and [JSON](#). In practice, the Web service typically provides an [object-oriented](#) Web-based interface to a database server, utilized for example by another Web server, or by a [mobile application](#), that provides a user interface to the end user. Another common application offered to the end user may be a [mashup](#), where a Web server consumes several Web services at different machines, and compiles the content into one user interface.

Ref: https://en.wikipedia.org/wiki/Web_service





Accedere ai contenuti

- API Sandbox
 - Permettono di testare le chiamate alle API prima di usarle nei propri programmi
 - Per vedere i contenuti di una pagina
 - Per vedere tutti gli articoli di una categoria

https://www.mediawiki.org/wiki/API:Main_page





Accedere ai contenuti

[https://it.wikipedia.org/wiki/Speciale:ApiSandbox#
action=query&
format=json&
prop=revisions&
titles=Istituto tecnico industriale statale Galileo Galilei \(Livorno\)&
rvprop=content](https://it.wikipedia.org/w/index.php?title=Speciale:ApiSandbox&action=query&format=json&prop=revisions&titles=Istituto_tecnico_industriale_statale_Galileo_Galilei_(Livorno)&rvprop=content)





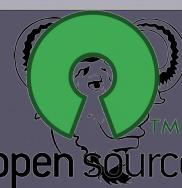
Accedere ai contenuti

Pagina di prova API

Utilizza questa pagina per fare pratica con le [API web service MediaWiki](#). Per ulteriori dettagli di utilizzo delle API, fai riferimento alla [documentazione API](#). Esempio: [ottenere il contenuto della pagina principale](#). Seleziona un'azione per vedere altri esempi. Nota che, anche se questa è una pagina per le prove, le azioni che esegui qui potrebbero modificare il wiki.

Espandi pannello Inoltra richiesta Pulisci

main	exportnowrap
action=query	Return the export XML without wrapping it in an XML result (same format as Special:Export). Can only be used with export.
prop=revisions	<input checked="" type="checkbox"/> iwurl
format=json	Whether to get the full URL if the title is an interwiki link.
Risultati	continue
	Quando più risultati sono disponibili, usa questo per continuare.
	<input checked="" type="checkbox"/> rawcontinue
	Return raw query - continue data for continuation.
	titles
	Istituto tecnico industriale statale Galileo Galilei (Livorno) <input type="button" value="X"/>
	Un elenco di titoli su cui lavorare. Il numero massimo di valori è 50 (500 per i bot).





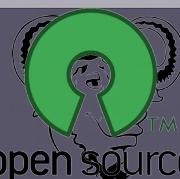
Accedere ai contenuti

Pagina di prova API

Utilizza questa pagina per fare pratica con le **API web service MediaWiki**. Per ulteriori dettagli di utilizzo delle API, fai riferimento alla [documentazione API](#). Esempio: [ottenere il contenuto della pagina principale](#). Seleziona un'azione per vedere altri esempi. Nota che, anche se questa è una pagina per le prove, le azioni che esegui qui potrebbero modificare il wiki.

Espandi pannello Inoltra richiesta Pulisci

main	URL di richiesta:	/w/api.php?action=query&format=json&prop=revi
action=query		
prop=revisions		
format=json		
Risultati	<pre>{ "batchcomplete": "", "query": { "pages": { "5090700": { "pageid": 5090700, "ns": 0, "title": "Istituto tecnico industriale statale Galileo Galilei (Livorno)", "revisions": [{ "contentformat": "text/x-wiki", "contentmodel": "wikitext", "*": " {{S scuole Livorno}}\n{{Scuola\n Nome = Istituto tecnico industriale statale Galileo Galilei\n Soprannome = \n Logo = \n Dimensione logo = \n Immagine = \n Dimensione immagine = \n Nazione = ITA\n citt\u00e0 = Livorno\n citt\u00e0link = Livorno\n Indirizzo = Via Galilei Galileo, 66/68\n Succursali = \n Tipo = Istituto tecnico industriale\n Ordinamento = pubblico\n Fondazione = 1887\n Preside = \n Dipendenti = \n Data = \n Nota = \n Studenti = \n Data2 = \n Nota2 = \n Testata = \n Motto =</pre>	





Pagine in una categoria

[https://it.wikipedia.org/wiki/Special:ApiSandbox#
action=query&
format=json&
list=categorymembers&
cmtitle=Categoria:Scuole_a_Livorno](https://it.wikipedia.org/wiki/Special:ApiSandbox#action=query&format=json&list=categorymembers&cmtitle=Categoria:Scuole_a_Livorno)





Pagine in una categoria

Pagina di prova API

Utilizza questa pagina per fare pratica con le **API web service MediaWiki**. Per ulteriori dettagli di utilizzo delle API, fai riferimento alla [documentazione API](#). Esempio: [ottenere il contenuto della pagina principale](#). Seleziona un'azione per vedere altri esempi. Nota che, anche se questa è una pagina per le prove, le azioni che esegui qui potrebbero modificare il wiki.

[Espandi pannello](#) [Inoltra richiesta](#) [Pulisci](#)

main	list=categorymembers
action=query	Elenca tutte le pagine in una categoria indicata.
list=categorymem...	Collegamenti alla guida Esempi
format=json	cmttitle <input type="text" value="Categoria:Scuole_a_Livorno"/> <input checked="" type="checkbox"/>
Risultati	Which category to enumerate (required). Must include the Categoria: prefix. Cannot be used together with <i>cmpageid</i> .
cmpageid	<input type="text"/> <input type="button" value="-"/> <input type="button" value="+"/> <input type="checkbox"/>
	Page ID of the category to enumerate. Cannot be used together with <i>cmttitle</i> .





Pagine in una categoria

Pagina di prova API

Utilizza questa pagina per fare pratica con le **API web service MediaWiki**. Per ulteriori dettagli di utilizzo delle API, fai riferimento alla [documentazione API](#). Esempio: [ottenere il contenuto della pagina principale](#). Seleziona un'azione per vedere altri esempi. Nota che, anche se questa è una pagina per le prove, le azioni che esegui qui potrebbero modificare il wiki.

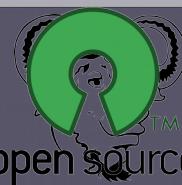
Espandi pannello Inoltra richiesta Pulisci

main
action=query
list=categorymembers
format=json

Risultati

URL di richiesta: /w/api.php?action=query&format=json&list=categorymembers

```
{  
    "batchcomplete": "",  
    "query": {  
        "categorymembers": [  
            {  
                "pageid": 3908112,  
                "ns": 0,  
                "title": "Istituto superiore di studi musicali Pietro  
Mascagni"  
            },  
            {  
                "pageid": 5090700,  
                "ns": 0,  
                "title": "Istituto tecnico industriale statale Galileo  
Galilei (Livorno)"  
            },  
            {  
                "pageid": 1953634,  
                "ns": 0,  
                "title": "Liceo Classico Niccolini Guerrazzi"  
            },  
            {  
                "pageid": 2058032,  
                "ns": 0,  
                "title": "Scuola Antonio Benci"  
            }  
        ]  
    }  
}
```





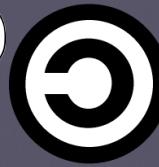
Progetto TEMDATA

	A	C	D	E	F	G	H
1	modello	lunghezza	larghezza	altezza	altezzasella	altezzaminima	altezzamax
2	Kawasaki ER-5 del 2006	2070	730	1070	800		
3	Kawasaki ER-6n del 2007	2100	760	1095	785		
4	Kawasaki GPZ 250						
5	Kawasaki GPZ 550 (ZX 550 A2)	2205	720	1245	780		
6	Kawasaki GTR 1000 (1986)	2290	780	1415	790		
7	Kawasaki H1 - Mach III del 1969	2095	1080	1150	810	120	
8	Kawasaki Zephyr(*)						
9	Kawasaki H1R 500 del 1969	2000	600	1140	720	145	
10	Kawasaki H2 - Mach IV del 1971	2080	850	1145		170	
11	Kawasaki Ninja H2(*)						
12	Kawasaki KLE 500 2007	2215	880	1270	850		
13	Kawasaki KLR 250 2005	2140	855	1210	855	270	
14	Kawasaki KLX 110				650	190	
15	Kawasaki KR 250 1977 ("KR 350 1978")					108	
16	Kawasaki KR 500 1981						
17	Kawasaki KR 750 1976						
18	Kawasaki KR-1 250 (S) [R] 1989	2015	690	1115	750		
19	Kawasaki KX 65 del 2008	1590	760	955	760		305
21	Kawasaki KXE 50 R (Automix)	2000	790		930		
22	Kawasaki Ninja 150 1996-2005	1975	719	1090		145	
23	Kawasaki Ninja 250R dal 1989 al 2007				745	155	
24	Kawasaki Ninja 300	2015	715	1110	785	140	
25	Kawasaki Ninja 400R	2100	760	1200	790	145	
26	KL 650 V						



Sabato 22 Ottobre 2016

F. Carrai - Gruppo Utenti Linux Livorno



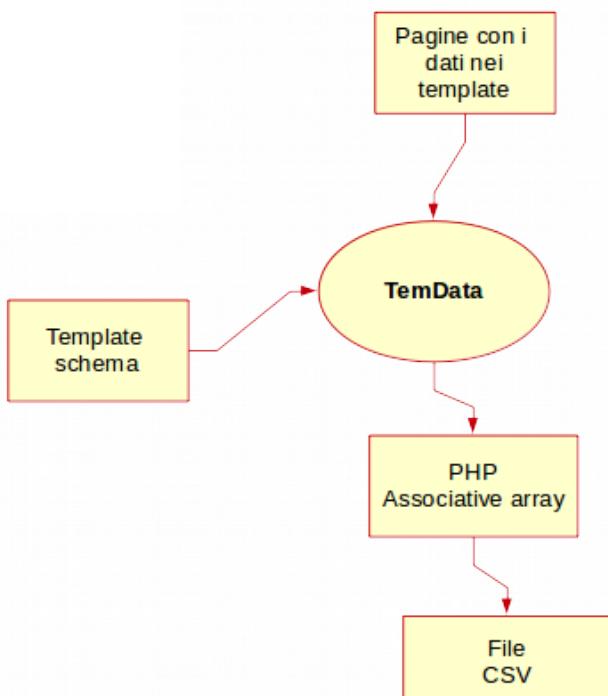
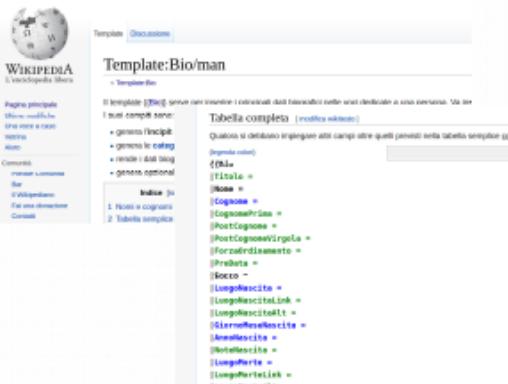
Linux Day 2016



Progetto TEMDATA

TemData - Data Flow Diagram

<https://it.wikipedia.org/wiki/Template:Bio/man>



```

{Bio
|Nome = Alfred Bernhard
|Cognome = Nobel
|PreData = pronuncia [{lingua svedese}|svedese]] {{IPA|['alfred 'bæ:ɳhɑd nɔ:bɛl]}}<ref>{{cita libro| cognome=Hedelin | nome=Per| titolo=Svenska uttals-lexikon | citta=Stoccolma | editore=Norstedts | anno=1997}}</ref> {{Link audio|Sv-Alfred Nobel.ogg|<small>ascolta</small>}}
|Sesso = M
|LuogoNascita = Stoccolma
|GiornoMeseNascita = 21 ottobre
|AnnoNascita = 1833
|LuogoMorte = Sanremo
|GiornoMeseMorte = 10 dicembre
|AnnoMorte = 1896
|Attività = chimico
|Attività2 = filantropo
|Nazionalità = svedese
|Immagine = AlfredNobel cropped.jpg
}}

```

```
Bio["Nome"]="Alfred Bernhard"
Bio["Cognome"]="Nobel"
Bio["Epoca"] = " "
...
Bio["Immagine"]=AlfredNobel_cropped.jpg"
```

Sabato 22 Ottobre 2016

E. Carrai - Gruppo Utenti Linux Liverne

15





Progetto TEMDATA

- Estrarre i **nomi dei campi** dal template (schema)
- In ogni pagina da processare **identificare lo schema**
- **Estrarre i nomi e valori dei campi**
(Nota: quelli che sono stati definiti!)
- Ripetere per tutte le pagine di interesse, es. quelle di una categoria
- Salvare in un file di testo, formato CSV





Parsing del testo

- Brutto ma buono: «**grep**»
- Potentissimo strumento per la ricerca ed estrazione di testi da una stringa
- Stringa di controllo delle azioni «illeggibile»
 - "`\A|(.*)=/s`"
 - "`\As\|.*=(.*?)/mU`"
- Tool online per studiare e provare le regular expression:

<https://regex101.com>





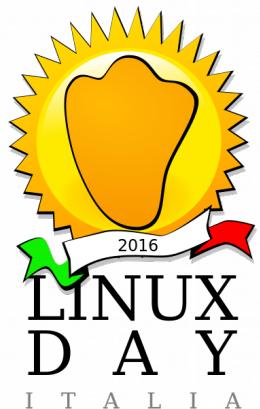
Progetto TEMDATA

Left Spreadsheet (Kawasaki Models):

	modello
1	Kawasaki ER-5 del 2006
2	Kawasaki ER-6n del 2007
3	Kawasaki GPZ 250
4	Kawasaki GPZ 550 (ZX 550 A2)
5	Kawasaki GTR 1000 (1986)
6	Kawasaki H1 - Mach III del 1969
7	Kawasaki Zephyr(*)
8	Kawasaki H1R 500 del 1969
9	Kawasaki H2 - Mach IV del 1971
10	Kawasaki Ninja H2(*)
11	Kawasaki KLE 500 2007
12	Kawasaki KLR 250 2005
13	Kawasaki KLX 110
14	Kawasaki KR 250 1977 ("KR 350 1978")
15	Kawasaki KR 500 1981
16	Kawasaki KR 750 1976
17	Kawasaki KR-1 250 (S) [R] 1989
18	Kawasaki KX 65 del 2008
19	Kawasaki KXE 50 R (Automix)
20	Kawasaki Ninja 150 1996-2005
21	Kawasaki Ninja 250R dal 1989 al 2007
22	Kawasaki Ninja 300
23	Kawasaki Ninja 400R

Right Spreadsheet (Template-Bio_Nati a Livo):

	A	B	H	I	L	M	O	R	S
	Nome	Cognome	Sesso	LuogoNascita	GiornoMeseNascita	AnnoNascita	LuogoMorte	GiornoMeseMorte	AnnoMorte
1	Abramo	Basevi	M	Livorno	29 novembre	1818	Firenze	25 novembre	1885
2	Abramo	Ferraresi	M	Livorno	marzo	1822	Magenta	4 giugno	1859
3	Achille	Polti	M	Livorno	23 ottobre	1825	Colico	11 novembre	1899
4	Ada	Biagini	F	Livorno	8 giugno	1900	Genova	14 dicembre	1944
5	Ademaro	Nicoletti Altimari	M	Livorno	6 aprile	1903	Rodi	30 luglio	1940
6	Adolfo	Mangini	M	Livorno	24 settembre	1854	Livorno	2 febbraio	1929
7	Adolfo	Tommasi	M	Livorno		1851	Firenze		1933
8	Adriano	Chiaramida	M	Livorno	8 luglio	1941			
9	Adriano	Lemmi	M	Livorno	30 aprile	1822	Firenze	23 maggio	1906
10	Adriano	Lenzi	M	Livorno	12 luglio	1937			
11	Agide	Albertina	F	Livorno	24 agosto	1873	Firenze	15 dicembre	1960
12	Alberto	Palau	M	Livorno	15 novembre	1874	Torino	26 settembre	1944
13	Alberto	Bolaffi	M	Livorno					
14	Alberto	Chiesa	M	Livorno	25 febbraio	1988			
15	Alberto	Merciali	M	Livorno	9 giugno	1900			?
16	Alberto	Micheli Pellegrini	M	Livorno	28 dicembre	1870	Bibbiena	7 gennaio	1943
17									
18	Alberto	Rossi	M	Livorno	17 maggio	1966			
19	Alberto	Zampieri	M	Livorno		1903	Pisa		1992
20	Aldo	Bardelli	M	Livorno	4 aprile	1912	Bologna	15 novembre	1971
21	Aldo	Mieli	M	Livorno	4 dicembre	1879	Florida	16 febbraio	1950
22	Aldo	Montano	M	Livorno	18 novembre	1978			
23	Aldo	Montano	M	Livorno	23 novembre	1910	Livorno	2 settembre	1996
	Aldo	Nadi	M	Livorno	23 aprile	1899	Los Angeles	10 novembre	1965



Sabato 22 Ottobre 2016

F. Carrai - Gruppo Utenti Linux Livorno





Visualizzare e processare i dati

- LibreOffice “Calc”



- Importare in database:



- ..e molte altre applicazioni libere!





TEMDATA Online

- Sorgenti presto disponibili on-line

<https://github.com/gulli-livorno>

The screenshot shows the GitHub profile page for the user 'gulli-livorno'. The profile picture is a circular icon featuring a penguin standing next to a server tower, with the text 'G.U.L.Li' at the bottom. The profile name is 'Gruppo Utenti Linux Livorno' and the handle is 'gulli-livorno'. Below the profile picture is a 'Follow' button. The GitHub header includes a search bar, 'Pull requests', 'Issues', and 'Gist' buttons. The main content area shows an 'Overview' tab selected, showing 'Repositories 3', 'Stars 0', and 'Followers 0'. Below this, there's a section for 'Popular repositories' with two listed:

- controllo_accessi_gulli**: Progetto Arduino per il controllo accessi attraverso una varco con moduli sonar HC-SR04. Rating: ★ 3, Language: Arduino.
- noobs**: Forked from [raspberrypi/noobs](#). Description: NOOBS (New Out Of Box Software) - An easy Operating System install manager for the Raspberry Pi. Rating: ★ 0, Language: Makefile.

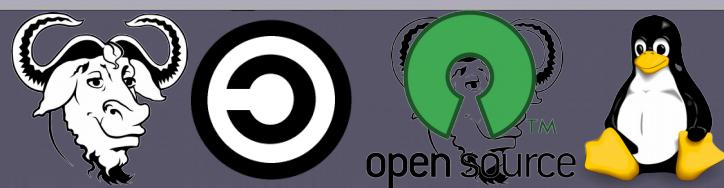
At the bottom, it says '3 contributions in the last year'.



Sabato 22 Ottobre 2016

F. Carrai - Gruppo Utenti Linux Livorno

20





Seguici su



Web

: <http://www.linux.livorno.it>



Facebook

: <https://www.facebook.com/luggulli>



Twitter

: https://twitter.com/gulli_linux



Google+

: <https://plus.google.com/+GULLILUG/>



Mailing list

: <http://lists.linux.it/listinfo/gulli>



Telegram

: https://telegram.me/gulli_linux

