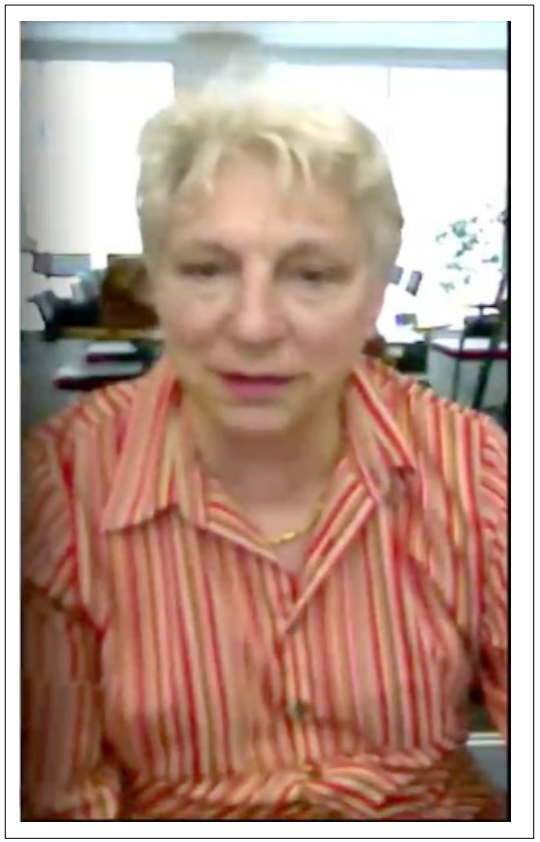
WIKIPEDIA'S STRUCTURED DATA CHALLENGE

ERIK MOELLER
TREVOR PARSCAL
SEMTECH CONFERENCE, JUNE 25, 2010

WIKIMEDIA FOUNDATION

PART 1:
OF HUMANS AND WIKITEXT
(AND TEMPLATES)



Editing Manchester Small-Scale Experimental Machine - Wikipedia, the free encyclopedia

http://en.wikipedia.org/w/index.php?title=Manchester_Small-Scale_Experimental_M

Google Gmail Hotmail Y! Mail Wave Identi.ca Twitter Voicemail Calendar Blogger Localhost Wikimedia Stuff Etherpads

Editing Manchester Small-Scale E...

Try Beta Log in / create account

article discussion edit this page history

Editing Manchester Small-Scale Experimental Machine

From Wikipedia, the free encyclopedia

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

search

Go Search

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox
- What links here
- Related changes
- Upload file
- Special pages

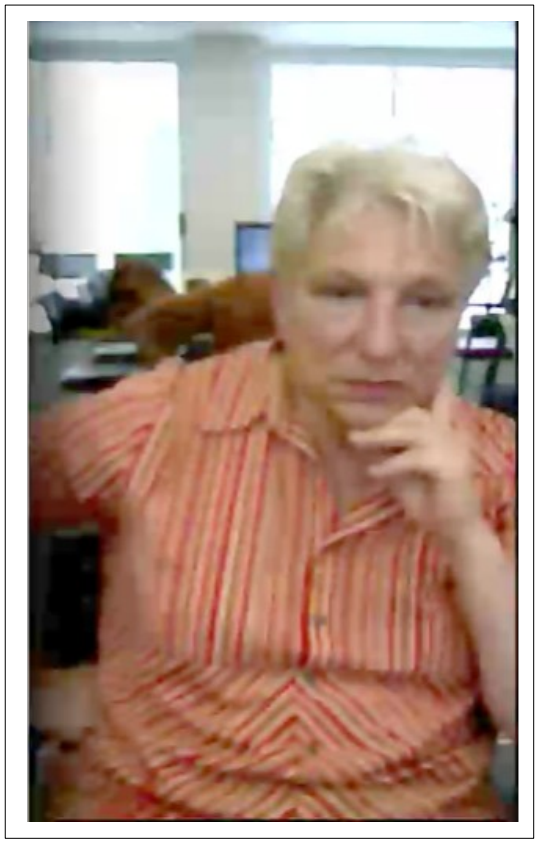
```
[[File:SSEM Manchester museum.jpg|thumb|300px|alt=A series of seven tall metal racks filled with
electronic equipment standing in front of a brick wall. Signs above each rack describe the functions
carried out by the electronics they contain.  Three visitors read from information stands to the left
of the image.|Replica of the Small-Scale Experimental Machine (SSEM) at the [[Museum of Science and
Industry (Manchester)|Museum of Science and Industry]] in [[Manchester]]]]
The '''Manchester Small-Scale Experimental Machine''' ('''SSEM'''), nicknamed '''Baby''', was the
world's first [[Von Neumann architecture|stored-program]] [[computer]]. It was built at the [[Victoria
University of Manchester]] by [[Frederic Calland Williams|Frederic C. Williams]], [[Tom Kilburn]] and
Geoff Tootill, and ran its first program on 21 June 1948.<ref>{{citation |last=Enticknap
|first=Nicholas |title=Computing's Golden Jubilee |journal=Resurrection |issue=20 |publisher=The
Computer Conservation Society |date=Summer 1998 |url=http://www.cs.man.ac.uk/CCS/res/res20.htm#d
|issn=0958-7403 |accessdate=19 April 2008.}}</ref>

The machine was not intended to be a practical computer but was instead designed as a [[testbed]] for
the [[Williams tube]], an early form of computer memory. Although considered "small and primitive" by
the standards of its time, it was the first working machine to contain all of the elements essential
to a modern electronic computer.<ref name=EarlyComputers /> As soon as the SSEM had demonstrated the
feasibility of its design, a project was initiated at the University to develop it into a more usable
computer, the [[Manchester Mark 1]]. The Mark 1 in turn quickly became the prototype for the
[[Ferranti Mark 1]], the world's first commercially available general-purpose computer.<ref
name=NapperMK1>{{citation |last=Napper |first=R. B. E. |title=Introduction to the Mark 1
|url=http://www.computer50.org/mark1/marklintro.html |publisher=The University of Manchester
|accessdate=4 November 2008.}}</ref>

The SSEM had a 32-[[bit]] [[word (computing)|word]] length and a [[computer memory|memory]] of
32 words. As it was designed to be the simplest possible stored-program computer, the only
```

Content that violates any copyrights will be deleted. Encyclopedic content must be **verifiable**. You irrevocably agree to release your contributions under the CC-BY-SA 3.0 License and the GFDL. You agree that a hyperlink or URL is sufficient attribution under the Creative Commons license. See the **Terms of Use** for details.

Edit summary (Briefly describe the changes you have made)

Save page Show preview Show changes Cancel | Editing help (opens in new window)

Done 1024x768



'''[[Wikitext]]''',
"it's kinda' messy"{{citation needed}}

Roles of contribution

- Descriptive markup (content)
  - Facts, figures, spelling and grammar fixes, etc.
  - Moderate expertise in Wikitext required
- Presentation markup (html, css)
  - Placement and styling of tables, images
  - Moderate expertise in HTML/CSS required
- Procedural markup (templates)
  - Creating info-boxes, citations, notices, etc.
  - Significant expertise in Wikitext required

Description, Presentation and Procedure (concept)

```
== Markup Language ==

A markup language is a modern
system for [[Annotation|
annotating]] a text in a way
that is syntactically
distinguishable from that
text.

Examples of markup languages
include:

* SGML, XML and HTML
* TeX and LaTeX
* Wikitext
```

Markup Language [edit]

A markup language is a modern system for annotating a text in a way that is syntactically distinguishable from that text
Examples of markup languages include:

- SGML, XML and HTML
- TeX and LaTeX
- Wikitext

Description, Presentation and Procedure (reality)

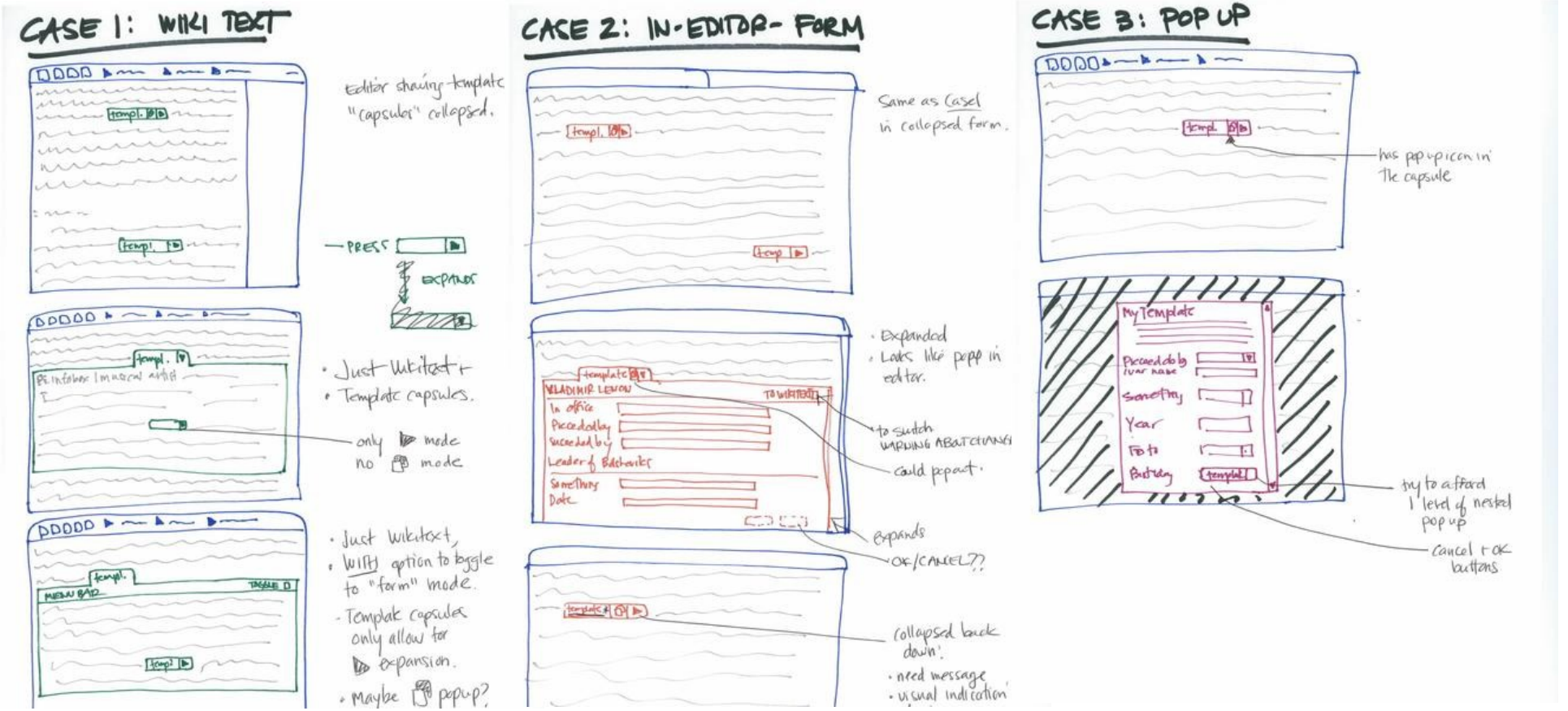
```
<!--BANNER ACROSS TOP OF PAGE-->
{| id="mp-topbanner" style="width:100%;
background:#f6f6f6; margin-top:1.2em; border:1px
solid #ccc;"
| style="width:61%; color:#000;" |
<!--"WELCOME TO WIKIPEDIA" AND ARTICLE COUNT-->
{| style="width:280px; border:none;
background:none;"
| style="width:280px; text-align:center; white-
space:nowrap; color:#000;" |
<div style="font-size:162%; border:none;
margin:0; padding:.1em; color:#000;">Welcome to
[[Wikipedia]],</div>
<div style="top:+0.2em; font-size:95%;">the
[[free content|free]] [[encyclopedia]] that
[[Wikipedia:Introduction|
anyone can edit]].</div>
<div id="articlecount" style="width:100%; text-
align:center; font-
size:85%;">[[Special:Statistics|
{{NUMBEROFARTICLES}}]] articles in [[English
language|English]]</div>
|}
```

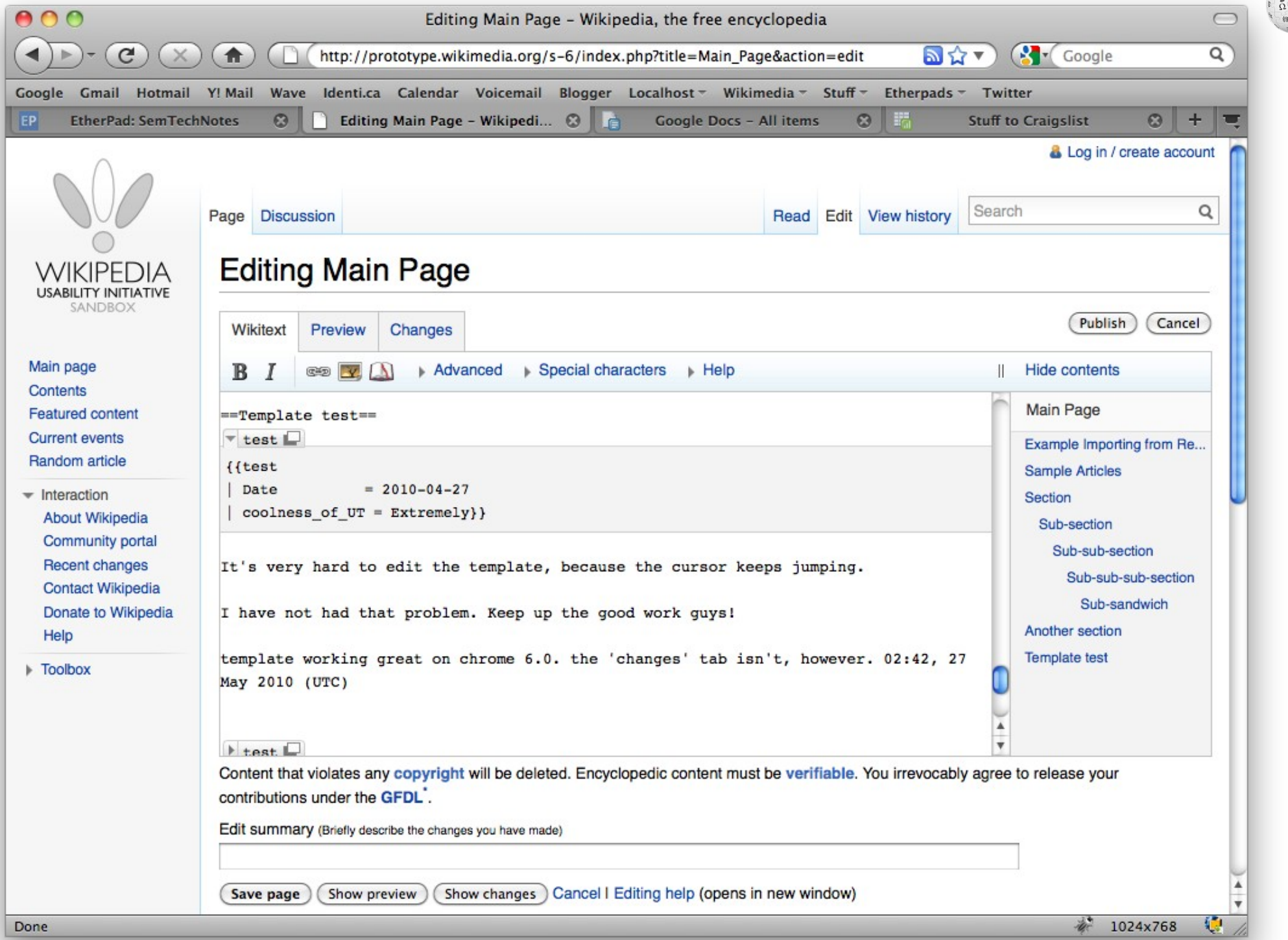
Welcome to Wikipedia,
the free encyclopedia that anyone can edit.
3,331,743 articles in English

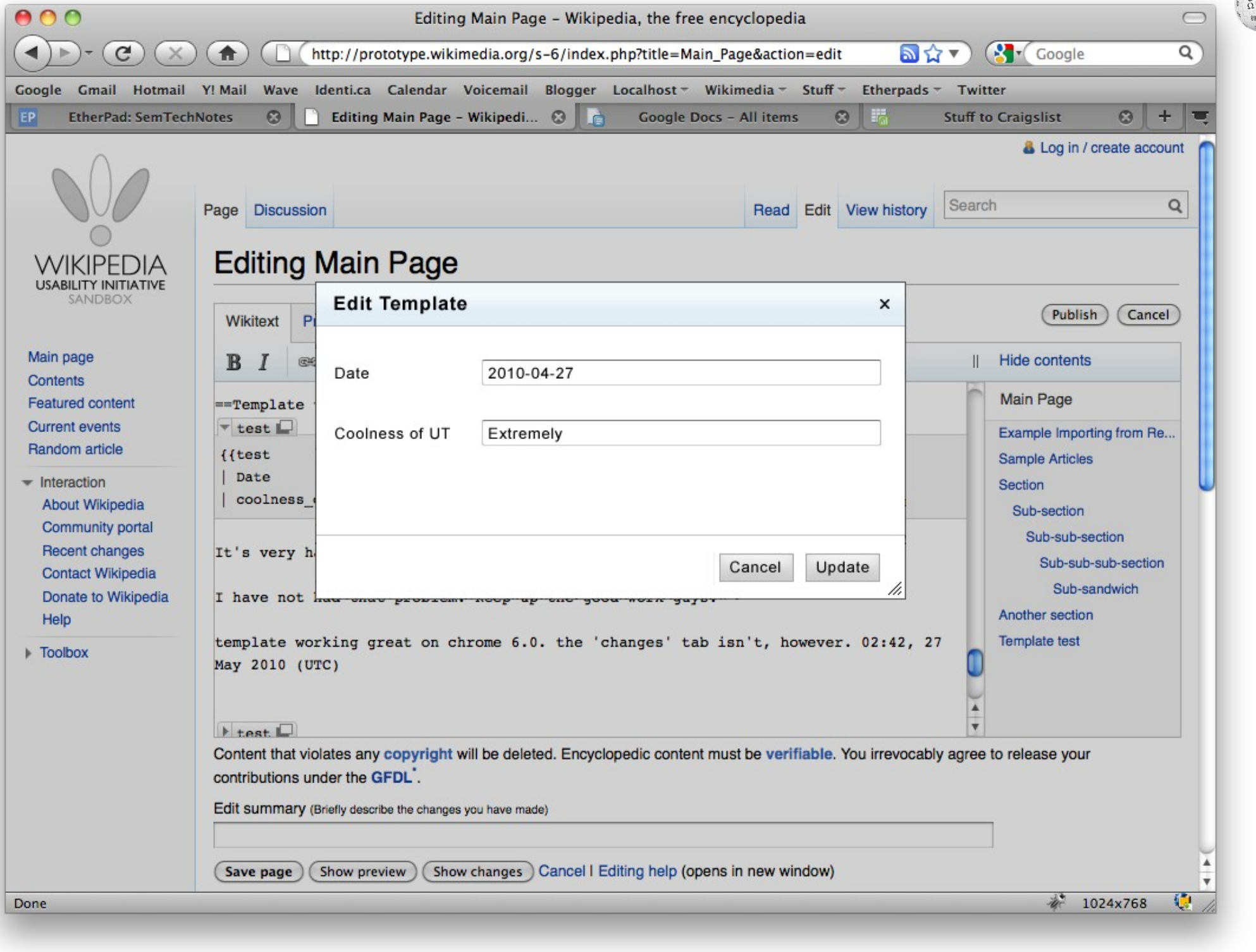
Why is visual editing so hard?

- Commingling
  - Description, presentation and procedural information are mixed together
- Ambiguity
  - Multiple styles of syntax can result in the same HTML output
  - Parsing doesn't happen semantically - we don't know what is creating what where and how, it's just a macro expander and a pile of regular expressions

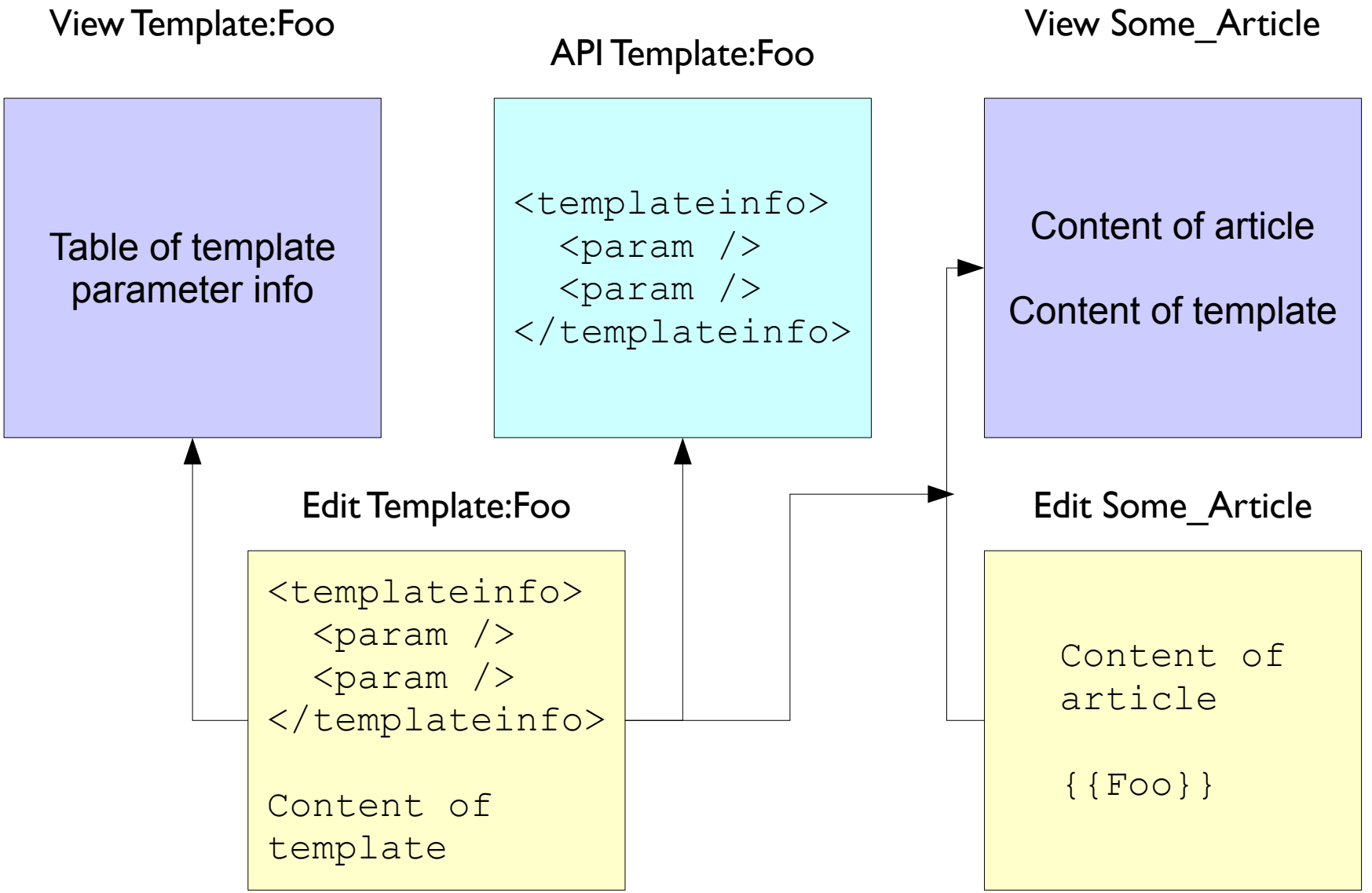
Interaction Methods




Editing Main Page - Wikipedia, the free encyclopedia
http://prototype.wikimedia.org/s-6/index.php?title=Main_Page&action=edit
Google
Gmail Hotmail Y! Mail Wave Identi.ca Calendar Voicemail Blogger Localhost Wikimedia Stuff Etherpads Twitter
EtherPad: SemTechNotes
Editing Main Page - Wikipedi...
Google Docs - All items
Stuff to Craigslist
Log in / create account
WIKIPEDIA
USABILITY INITIATIVE
SANDBOX
Page Discussion
Read Edit View history
Search
Editing Main Page
Wikitext Preview Changes
Publish Cancel
Advanced Special characters Help
Hide contents
Main page
Contents
Featured content
Current events
Random article
Interaction
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Donate to Wikipedia
Help
Toolbox
==Template test==
test
{{test
| Date = 2010-04-27
| coolness_of_UT = Extremely}}
It's very hard to edit the template, because the cursor keeps jumping.
I have not had that problem. Keep up the good work guys!
template working great on chrome 6.0. the 'changes' tab isn't, however. 02:42, 27 May 2010 (UTC)
test
Main Page
Example Importing from Re...
Sample Articles
Section
Sub-section
Sub-sub-section
Sub-sub-sub-section
Sub-sandwich
Another section
Template test
Content that violates any copyright will be deleted. Encyclopedic content must be verifiable. You irrevocably agree to release your contributions under the GFDL.
Edit summary (Briefly describe the changes you have made)
Save page Show preview Show changes Cancel | Editing help (opens in new window)
Done
1024x768


Editing Main Page – Wikipedia, the free encyclopedia
http://prototype.wikimedia.org/s-6/index.php?title=Main_Page&action=edit
Google
Gmail Hotmail Y! Mail Wave Identi.ca Calendar Voicemail Blogger Localhost Wikimedia Stuff Etherpads Twitter
EtherPad: SemTechNotes
Editing Main Page – Wikipedi...
Google Docs – All items
Stuff to Craigslist
Log in / create account
WIKIPEDIA
USABILITY INITIATIVE
SANDBOX
Page Discussion
Read Edit View history
Search
Editing Main Page
Edit Template
Date
2010-04-27
Coolness of UT
Extremely
Cancel
Update
Wikitext
Publish
Cancel
Hide contents
Main Page
Example Importing from Re...
Sample Articles
Section
Sub-section
Sub-sub-section
Sub-sub-sub-section
Sub-sandwich
Another section
Template test
Main page
Contents
Featured content
Current events
Random article
Interaction
About Wikipedia
Community portal
Recent changes
Contact Wikipedia
Donate to Wikipedia
Help
Toolbox
==Template
test
{{test
| Date
| coolness_
It's very h
I have not
template working great on chrome 6.0. the 'changes' tab isn't, however. 02:42, 27 May 2010 (UTC)
test
Content that violates any copyright will be deleted. Encyclopedic content must be verifiable. You irrevocably agree to release your contributions under the GFDL.
Edit summary (Briefly describe the changes you have made)
Save page
Show preview
Show changes
Cancel | Editing help (opens in new window)
Done
1024x768

Template Info Extension



Beyond Templates

Interlanguage links

[[af:Kreasionisme]]
[[ar:قلخالا ةيرظن]]
[[az:Kreasionizm]]
[[bg:Креационизъм]]
[[ca:Creacionisme]]
[[cs:Kreacionismus]]
[[da:Kreationisme]]
[[de:Kreationismus]]

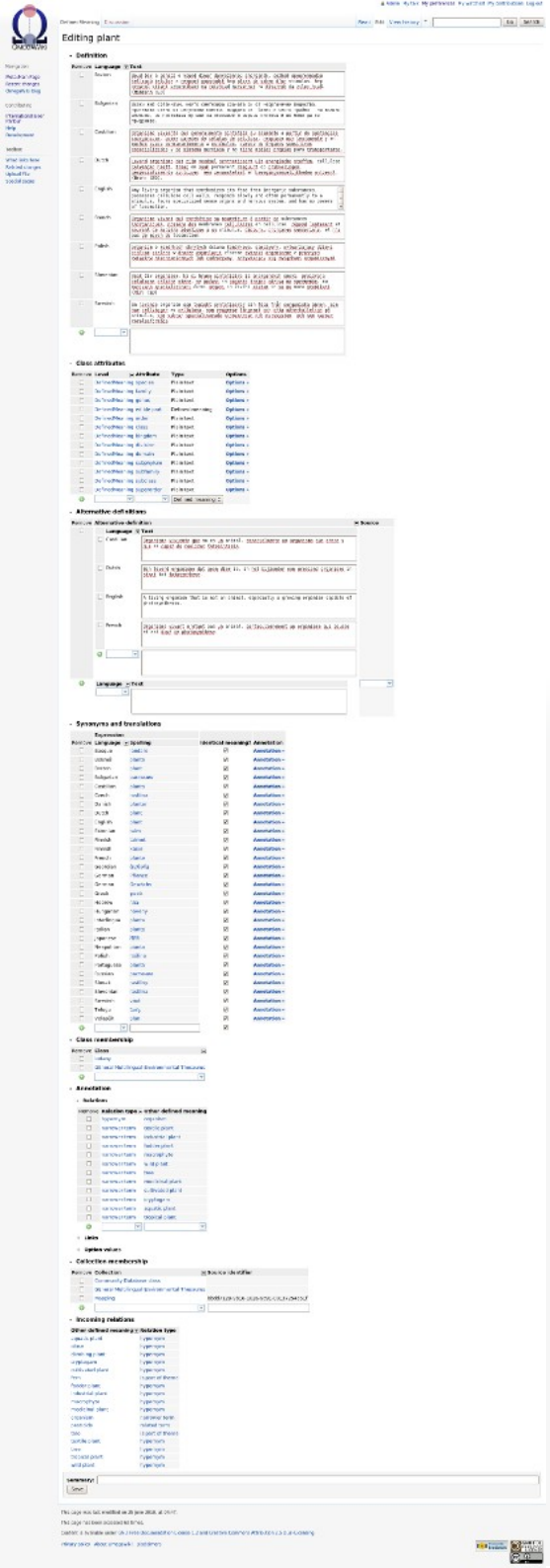
Categories

[[Category:Creationism]]
[[Category:Origin of life]]
[[Category:Theism]]
[[Category:Theology]]
[[Category:Christian terms]]
[[Category:Creation myths]]

Citations

{{citation
|date=2004
|author=[[Eugenie Scott|Eugenie C. Scott]] (with forward by Niles Eldredge)
|title=Evolution vs. Creationism: An Introduction
|place=Berkley & Los Angeles, California
|publisher=University of California Press
|page=114
|url=http://books.google.com/books?id=03b_a0monNYC&printsec=frontcover&dq=evolution+vs.+creationism&hl=en&ei=k1EZTMTRD86LkAWu2-1C&sa=X&oi=book_result&ct=result&resnum=1&ved=0CC4Q6AEwAA#v=onepage&q&f=false |isbn=0-520-24650-0 |accessdate=16 June 2010
}}

PART 2:
WIKI DATA NOW!



The Multilingual Ontology: OmegaWiki

OMEGAWIKI

Navigation

Meta:Main Page
Recent changes
OmegaWiki blog

Contributing

International Beer Parlour
Help
Development

Toolbox

What links here
Related changes
Upload File
Special pages

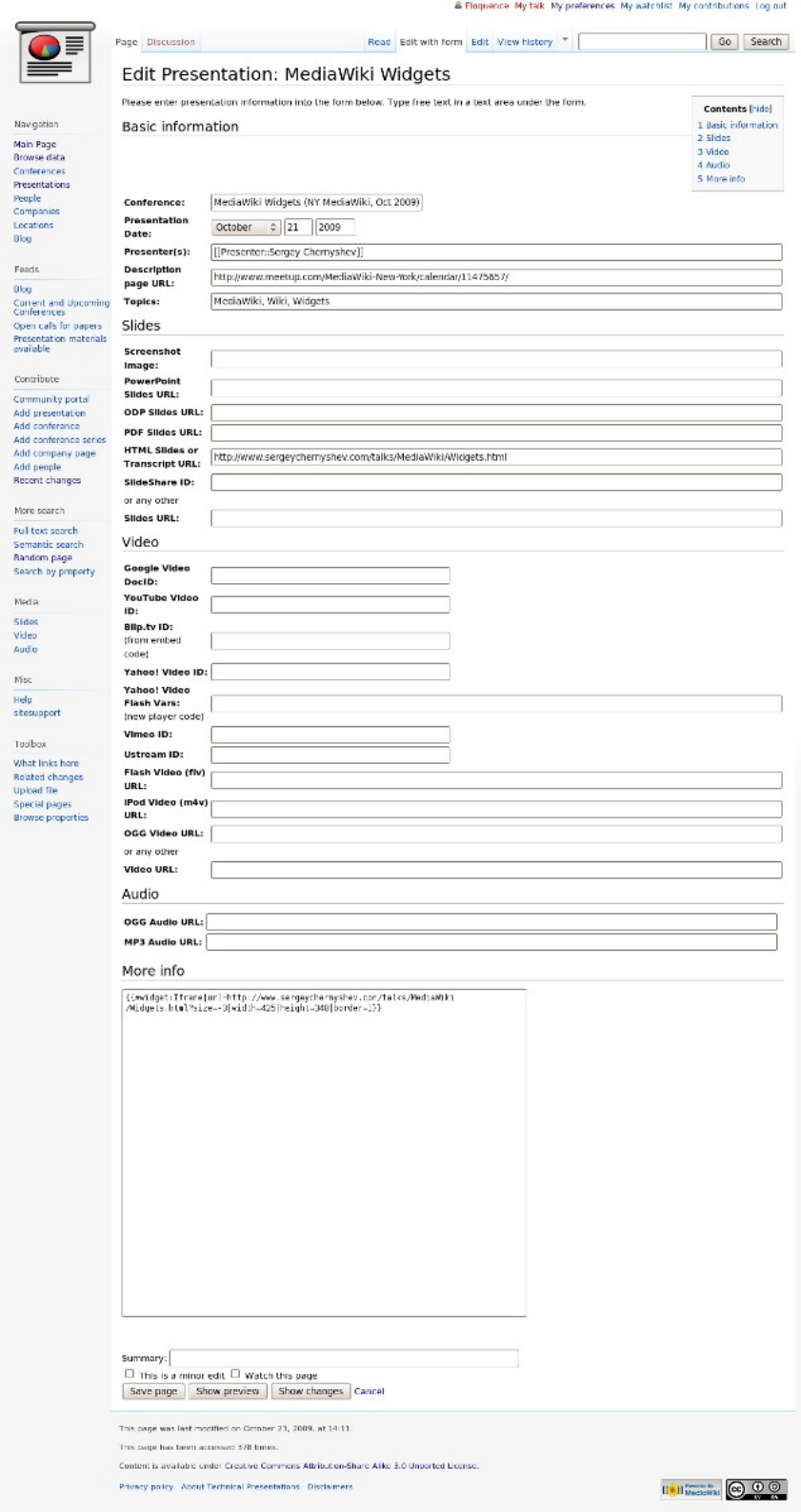
DefinedMeaning | Discussion

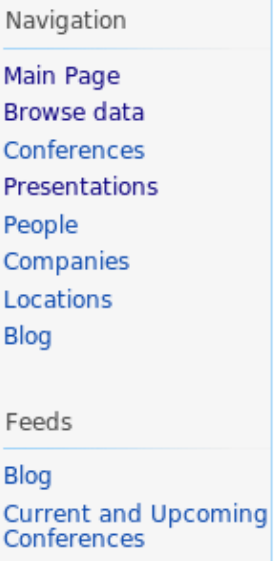
Editing plant

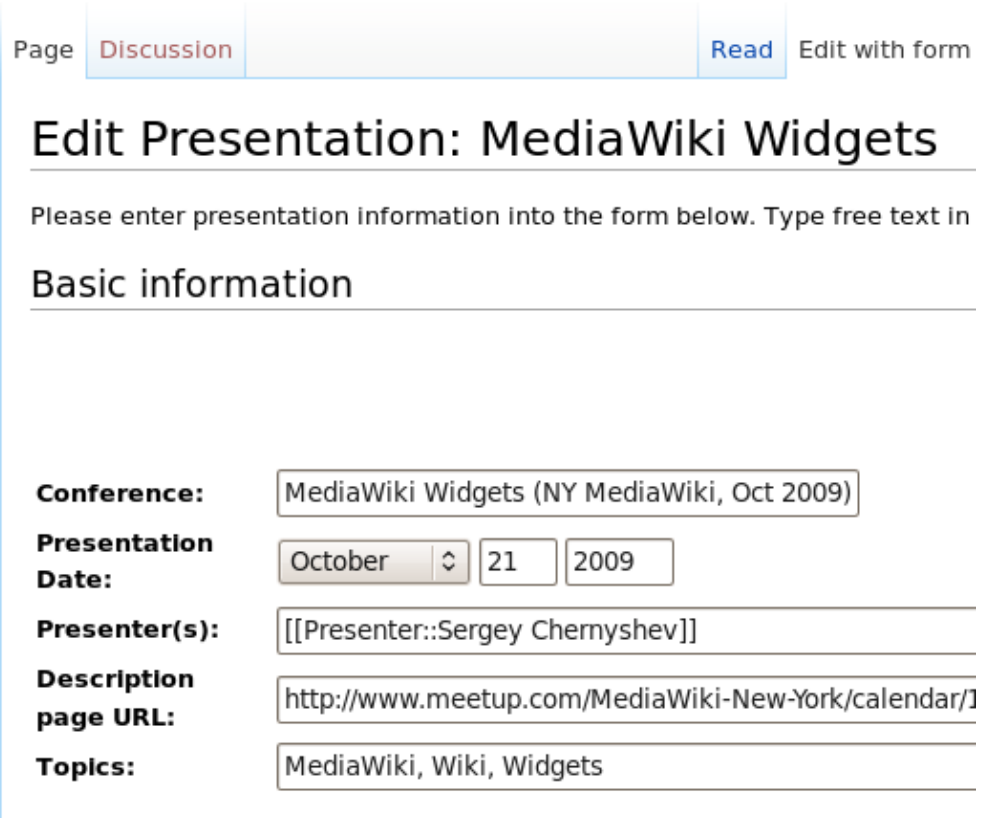
- Definition

Remove	Language	Text
☐	Breton	Boud bev a genaoz e voued diwar danvezennoù anorganek, dezhañ spewrennadoù kelligoù keluloz a respont goustadik hag alies da viken d'ur stimulus, hep organoù skiant arbennikaet na reizhiad nervennek ha divarrek da zilec'hiañ. (Mammenn CED)
☐	Bulgarian	Всеки жив организъм, който синтезира храната си от неорганични вещества, притежава стени от целулозни клетки, поддава се бавно и често трайно на външни влияния, не притежава органи за осезание и нервна система и не може да се придвижва.
☐	Castilian	Organismo viviente que generalmente sintetiza su alimento a partir de sustancias inorgánicas, posee paredes de células de celulosa, responde muy lentamente y en muchos casos permanentemente a estímulos, carece de órganos sensitivos especializados y de sistema nervioso y no tiene medios propios para transportarse.
☐	Dutch	Levend organisme dat zijn voedsel synthetiseert uit anorgische stoffen, cellulose celwanden heeft, traag en vaak permanent reaguurt op prikkelingen, gespecialiseerde zintuigen, een zenuwstelsel en bewegingsmogelijkheden ontbeert. (Bron: CED).
☐	English	Any living organism that synthesizes its food from inorganic substances, possesses cellulose cell walls, responds slowly and often permanently to a stimulus, lacks specialized sense organs and nervous system, and has no powers of locomotion.

The Semantic Way: Semantic MediaWiki and Semantic Forms







Extraction:
DBPedia

OPENLINK SOFTWARE

About Stanley Kubrick

An Entity of Type : Person, within Data Space : dbpedia.org

Stanley Kubrick (July 26, 1928 – March 7, 1999) was an American director, writer, producer, and photographer of films, who lived in England during most of the last 40 years of his career. Kubrick was noted for the scrupulous care with which he chose his subjects, his slow method of working, the variety of genres he worked in, his technical perfectionism and his reclusiveness about his films and personal life.

Property	Value
dbpedia-owl:abstract	Stanley Kubrick (July 26, 1928 – March 7, 1999) was an American director, writer, producer, and noted for the scrupulous care with which he chose his subjects, his slow method of working, the personal life. He worked far beyond the confines of the Hollywood system, maintaining almost co himself, but with the rare advantage of big-studio financial support for all his endeavors. Oscar n effects in '. Kubrick is widely acknowledged as one of the most accomplished, innovative and influ controversial films that have often been perceived as a reflection of his obsessive and perfectionis later films often have elements of surrealism and expressionism eschewing structured linear nar cautious optimism when viewed more carefully. His works are noted as some of the "most origin »more»
dbpedia-owl:activeYearsEndYear	1999-01-01 00:00:00(xsd:date)
dbpedia-owl:activeYearsStartYear	1951-01-01 00:00:00(xsd:date)
dbpedia-owl:birthDate	1928-07-26(xsd:date)

Application: WikiPics

Wikipics 0.1α (experimental) Wikimedia Deutschland e.V.

German go Note: this is a thesaurus lookup, not a full text search. Multiple words are handeled as a single phrase. Only exact matches of complete phrases will be found.

Gnus: *Gnus sind afrikanische Antilopen, die in großen Herden leben und zur Gruppe der Kuhantilopen gehören.*

Wilderbeest
jpeg, 3615x2380, 2M
BY

Black-wildebeest-aka-gnu
jpeg, 652x388, 82K
PD

Blue wildebeest
jpeg, 773x566, 87K
CC-BY-SA, GFDL

Gnou Thoiry 19801
jpeg, 1984x1398, 2M
PD

Gnu zebre
jpeg, 736x488, 120K
CC-BY-SA, GFDL

Wildebeest mikumi
jpeg, 1600x1067, 373K
GFDL

Wildebeest, Amboseli National Park
jpeg, 2451x1637, 1M
CC-BY-SA

Ngorongoro Wildebeest
jpeg, 1210x806, 678K
CC-BY-SA

Gnu 02
jpeg, 734x489, 112K
CC-BY-SA, GFDL

Gnusmall
jpeg, 461x307, 56K
CC-BY-SA, GFDL

Gnu head
jpeg, 640x480, 89K

Gnou 04 2005-07-01
jpeg, 1024x768, 161K
CC-BY-SA, GFDL

Weissbartgnus Zoo Leipzig
jpeg, 2844x1696, 2M
CC-BY-SA, GFDL

Troupeaudegnous-sigean
jpeg, 456x177, 11K
CC-BY-SA, GFDL

Gnu 01
jpeg, 728x488, 134K
CC-BY-SA, GFDL

Wildebeest Steve Evans
jpeg, 1080x720, 93K
BY

Gnou 07 2005-07-01
jpeg, 1024x768, 145K
CC-BY-SA, GFDL

Narrower: Streifengnu, Weißschwanzgnu
Related: Zebra, Afrika, Hornträger, Kenia, Prädator, Savanne, Serengeti, Tansania, Weißschwanzgnu
Broader: Fauna of East Africa, Mammals of Kenya, Megafauna of Africa, Fauna of Angola, Fauna of Namibia, Fauna of South Africa, Fauna of Tanzania, Fauna of Zambia, Mammals of Africa, Fauna of Africa, Kuhantilopen, Hornträger, Netzmagen
Wiki pages: de:Gnus, en:Wildebeest, es:Connochaetes, fr:Gnou, it:Connochaetes, nl:Gnoes, pl:Connochaetes, pt:Gnu

gnu German go Note: this is a thesaurus lookup, not a full text search. Multiple words are handeled as a single phrase. Only exact matches of complete phrases will be found.

GNU: *GNU ist das im Rahmen des GNU-Projekts in Entwicklung befindliche, vollständig freie Betriebssystem.*

En-gnu.ogg [more/details...]

Narrower: GNU Privacy Projekt, LibeRTOS, Minimal SYStem, XCSoar, UFRaw, CLISP, Data Display Debugger, GNAT Modified General Public License, GNAT Programming Studio, GNU-Projekt, GNU-Varianten, GNU Common Lisp, GNU Make, GNU PDF, GNU Pascal, GNU Toolchain, GnuWin32, NoteEdit, Texinfo, WinPT [more...]
Related: Red Hat, Richard Stallman, Software, Standard C Library, TeX, Ubuntu, Unicode, Unix, X.Org-Server, Yacc
Broader: Freies Betriebssystem, Microkernel-based operating systems, Computing acronyms, Free software operating systems, GNU project software, GNU-Projekt, Akronym, Freie Software, Linux-Software, Unix-Programm, Betriebssystem, Copyright, Unix [more...]

Gnus: *Gnus sind afrikanische Antilopen, die in großen Herden leben und zur Gruppe der Kuhantilopen gehören.*

[more/details...]

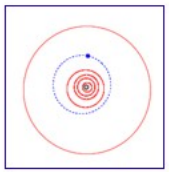
Narrower: Streifengnu, Weißschwanzgnu
Related: Zebra, Afrika, Hornträger, Kenia, Prädator, Savanne, Serengeti, Tansania, Weißschwanzgnu
Broader: Fauna of East Africa, Mammals of Kenya, Megafauna of Africa, Fauna of Angola, Fauna of Namibia, Fauna of South Africa, Fauna of Tanzania, Fauna of Zambia, Mammals of Africa, Fauna of Africa, Kuhantilopen, Hornträger, Netzmagen [more...]

GNU-Projekt: *Das GNU-Projekt wurde von Richard Stallman mit dem Ziel gegründet, ein vollständig freies Betriebssystem, genannt GNU, zu entwickeln.*

[more/details...]

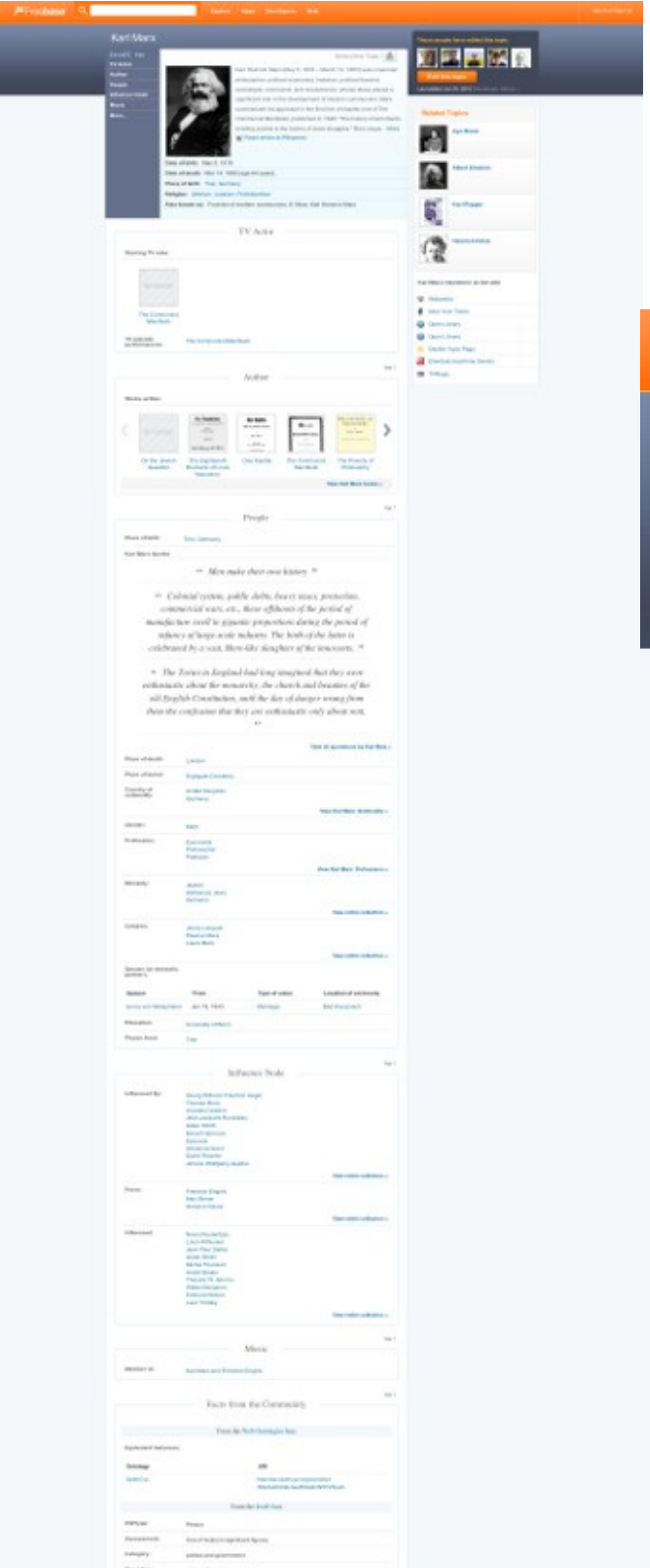
Narrower: Free software movement, GCIDE, GNU-Darwin, GNU.FREE, GNU Panama, GNU people, GNU project software, Gnits Standards, Gnulib, GNAT Modified General Public License, GNU Coding Standards, GNU Manifesto, GNU, GNUstep, GPL linking exception, Gpl-violations.org, GNU-Lizenz für freie Dokumentation, GNU Affero General Public License, GNU General Public License, GNU Lesser General Public License [more...]
Related: Linux (Kernel), Massachusetts Institute of Technology, Open Source, Rekursives Akronym, Richard Stallman, Unix, Unixoide Systeme
Broader: Copyleft media, Free software projects, GNU, Free Software Foundation, Urheberrecht [more...]

GNU (Asteroid): *(9965) GNU ist ein Asteroid des Hauptgürtels, der am 5. März 1992 im Rahmen des Spacewatch-Programms entdeckt wurde.*



[more/details...]

Related: Stallman (Asteroid), Torvalds (Asteroid), Spacewatch
Broader: C-type asteroids, Hauptgürtelasteroid, Asteroidengürtel [more...]



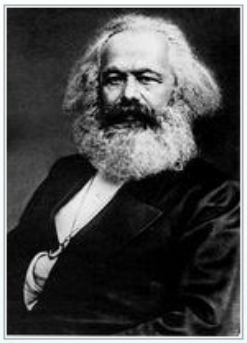
The Web 2.0 Way: Freebase

Freebase | Explore | Apps | Developers | Help

Karl Marx

Scroll to:
- TV Actor
- Author
- People
- Influence Node
- Music
- More...

Embed this Topic



Karl Heinrich Marx (May 5, 1818 – March 14, 1883) was a German philosopher, political economist, historian, political theorist, sociologist, communist, and revolutionary, whose ideas played a significant role in the development of modern communism. Marx summarized his approach in the first line of chapter one of The Communist Manifesto, published in 1848: "The history of all hitherto existing society is the history of class struggles." Marx argue... More

Read article at Wikipedia

Date of birth: May 5, 1818

Date of death: Mar 14, 1883 (age 64 years)

Place of birth: Trier, Germany

Religion: Atheism, Judaism, Protestantism

Also known as: Founder of modern communism, K. Marx, Karl Heinrich Marx

TV Actor

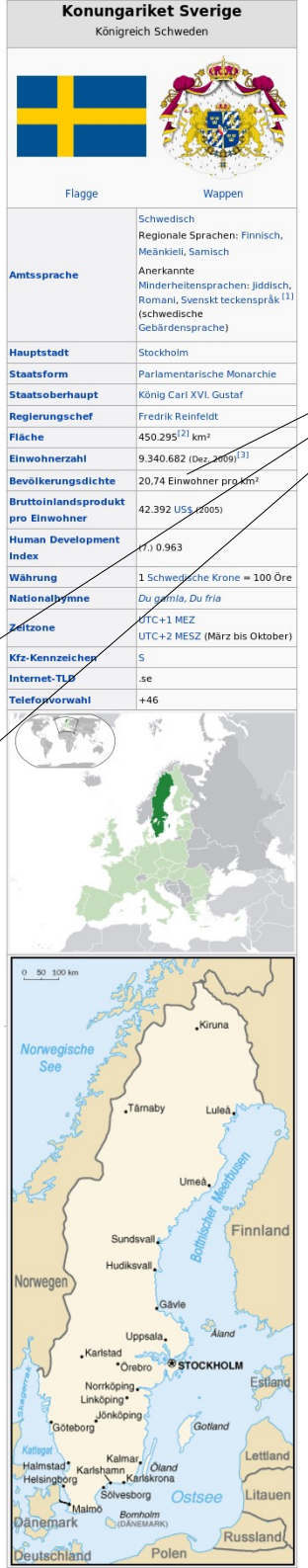
Starring TV roles

NO IMAGE

The Communist Manifesto

TV episode performances: The Communist Manifesto

PART 3:
YOUR MISSION
(SHOULD YOU DECIDE TO ACCEPT IT)



A Wikidata Commons

- Centralized repository
- Search and retrieval
  - Wikipedia list generation
- Fully multilingual
  - No monolingual strings
  - Support for locales
  - Bootstrap small Wikipedias
- Support for external data
- Rich APIs and exports
- Data/layout separation
  - Editable via forms
- Scales. And scales. And scales.

Will you help us build it?

LET'S TALK

erik@wikimedia.org
tparscal@wikimedia.org