

# Indywiki: Visually browsing Wikipedia



Markos Gogoulos  
mgogoulos@gmail.com  
FOSDEM 2008

# Why the project was started



*“Some times browsing over the plethora of information available through Wikipedia pages becomes a boring activity...”*

Indywiki project aims to explore different ways of browsing the Wikipedia universe!



# More specifically...

- Sometimes you just need to get a **quick idea** on an article
- get all images of an article **together** (in tenths)
- use as testbed for changes/experimentation
- ...and give something back to the amazing community of Wikipedia



# Critics

- Some people shout that *'Wikipedia is unreliable'*
- One should first check current research on Wikipedia reliability and then talk.
- I couldn't care less :) This is an open source project/experiment!




# Visually browsing Wikipedia?


Indywiki: exploring wikipedia visually


File View Help


```
def add5(x):
    return x+5


def dotwrite(ast):
    nodename = getNodeName()
    label=symbol.syn_name.get(int(ast[0]),ast[0])
    print "%s (%s)" % (nodename, label)
    if isinstance(ast[1], str):
        print "%s" % ast[1]
    else:
        print ""
    else:
        print ""
    print ""
    children = []
    for n, child in enumerate(ast[1:]):
        children.append(dotwrite(child))
    print "%s" % " ".join(children)
    for name in children:
        print "%s" % name,
```




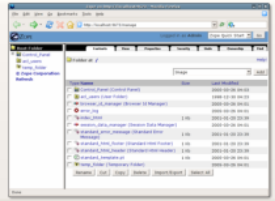


















**CONTENTS**

- [Python \(programming language\)](#)
- [History](#)
- [Future development](#)
- [Python 3000](#)
- [Philosophy](#)
- [Timeline and compatibility](#)
- [Features](#)
- [Usage](#)
- [Syntax and semantics](#)
- [Implementations](#)
- [Interpretational semantics](#)
- [Standard library](#)
- [Programming philosophy](#)
- [Neologisms](#)
- [Influences on other languages](#)
- [References](#)

Python is a general-purpose, high-level programming language. Its design philosophy emphasizes programmer productivity and code readability. Python's core syntax and semantics are minimalist, while the standard library is large and comprehensive.

Python supports multiple programming paradigms (primarily functional, object oriented and imperative), and features a fully dynamic type system and automatic memory management; it is thus similar to Perl, Ruby, Scheme, and Tcl.

Python was first released by Guido van Rossum in 1991. The language has an open, community-based development model managed by the non-profit Python Software Foundation. While various parts of the language have formal specifications and standards, the language as a whole is not formally specified. The de facto standard for the language is the CPython implementation.

**LINKS**

- [.NET framework](#)
- [.NET platform](#)
- [18 April](#)
- [2005](#)
- [2006](#)
- [2007](#)
- [21 April](#)
- [7 December](#)
- [7 March](#)
- [9 February](#)
- [ABC programming language](#)
- [ABC \(programming language\)](#)
- [ALGOL 68](#)
- [Academic conference](#)
- [Air Canada](#)
- [Alex Martelli](#)
- [Amoeba distributed operating system](#)

Indywiki search on <http://en.wikipedia.org>. Article [Python\\_\(programming language\)](#) has 1 images.

100%

“browsing  
an article ”



# Visually browsing Wikipedia?

Indywiki: exploring wikipedia visually

File View Help

def add5(x):  
 return x+5

def dotwrite(ast):  
 nodename = getNodeName()  
 label=symbol.sym\_name.get(int(ast[0]),ast[0])  
 print '%s [%s]' % (nodename, label),  
 if isinstance(ast[1], str):  
 if ast[1].strip():  
 print '%s' % ast[1]  
 else:  
 print ''  
 else:  
 print '['  
 children = []  
 for n, child in enumerate(ast[1:]):  
 children.append(dotwrite(child))  
 print ' %s -> {' % nodename,  
 for name in children:  
 print '%s' % name,

Ruby

PROGRAMMING REPUBLIC

@Groovy

Python add5 syntax.svg

```
def add5(x):  
    return x+5
```

```
def dotwrite(ast):  
    nodename = getNodeName()  
    label=symbol.sym_name.get(int(ast[0]),ast[0])  
    print '%s [%s]' % (nodename, label),  
    if isinstance(ast[1], str):  
        if ast[1].strip():  
            print '%s' % ast[1]  
        else:  
            print ''  
    else:  
        print '['  
        children = []  
        for n, child in enumerate(ast[1:]):  
            children.append(dotwrite(child))  
        print ' %s -> {' % nodename,  
        for name in children:  
            print '%s' % name,
```

Image link

Python was first released by Guido van Rossum in 1991. The language has an open, community-based development model managed by the non-profit Python Software Foundation. While various parts of the language have formal specifications and standards, the language as a whole is not formally specified. The de facto standard for the language is the CPython implementation.

Python (programming language)  
GO

CONTENTS

- [Python \(programming language\)](#)
- [History](#)
- [Future development](#)
- [Python 3000](#)
- [Philosophy](#)
- [Timeline and compatibility](#)
- [Features](#)
- [Usage](#)
- [Syntax and semantics](#)
- [Implementations](#)
- [Interpretational semantics](#)
- [Standard library](#)
- [Programming philosophy](#)
- [Neologisms](#)
- [Influences on other languages](#)
- [References](#)

LINKS

- [.NET framework](#)
- [.NET platform](#)
- [18 April 2005](#)
- [2006](#)
- [2007](#)
- [21 April 2007](#)
- [7 December 2007](#)
- [9 February 2007](#)
- [ABC programming language](#)
- [ABC \(programming language\)](#)
- [ALGOL 68](#)
- [Academic conference](#)
- [Air Canada](#)
- [Alex Martelli](#)
- [Amoeba distributed operating system](#)

Indywiki search on <http://en.wikipedia.org>. Article Python\_(programming\_language) has 1 images.

100%

“user clicks  
on a  
'native' icon”



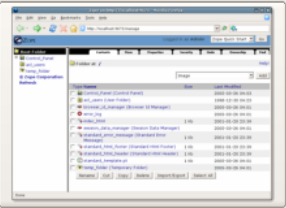
...where  
native is an  
icon found on  
the article




# Visually browsing Wikipedia?


Indywiki: exploring wikipedia visually

File View Help





```
def add5(x):  
    return x+5  
  
def dotwrite(ast):  
    nodename = getNodeName()  
    label=symbol.syn_name.get(int(ast[0]),ast[0])  
    print '% [label="%s" % (nodename, label),  
    if isinstance(ast[1], str):  
        if ast[1].strip():  
            print '% "%s"' % ast[1]  
        else:  
            print ""  
    else:  
        print ""  
        children = []  
        for n, child in enumerate(ast[1:]):  
            children.append(dotwrite(child))  
        print '% "%s"' % nodename,  
        for name in children:  
            print '% "%s"' % name,
```



**CONTENTS**  
[Zope](#)  
[History](#)  
[Technical features](#)  
[Zope 3](#)  
[Five and the Future of Zope](#)  
[Zope Page Templates](#)  
[References](#)

Zope is an open-source, object-oriented web application server written in the Python programming language. Zope stands for "Z Object Publishing Environment." It can be almost fully managed with a web-based user interface. Zope publishes on the web Python objects that are typically persisted in an object database, ZODB. Basic object types, such as documents, images, and page templates, are available for the user to create and manage through the web. Specialized object types, such as wikis, blogs, and photo galleries, are available as third-party add-ons (called products), and there is a thriving community of small businesses creating custom web applications as Zope products.

There are two major generations of the software in use today. As of July 2007, Zope 2.10.4 is the latest stable release of Zope 2 codebase, and Zope 3.3.1 is the latest release of Zope 3. Zope is distributed under the terms of the Zope Public License, a free software license.

**LINKS**  
[2004](#)  
[2007](#)  
[Content management](#)  
[ERP5](#)  
[Fredericksburg, Virginia](#)  
[Guido van Rossum](#)  
[HTML](#)  
[Hadar Pedhazur](#)  
[Information hiding](#)  
[Internationalization and localization](#)  
[January 14](#)  
[METAL](#)  
[November 6](#)  
[Nuxeo CPS](#)  
[Plone \(content management system\)](#)  
[Python \(programming](#)

Indywiki search on <http://en.wikipedia.org>. Article Zope has 1 images.

100%

user clicks  
on a  
'non\_native'  
image

...where  
non\_native  
is the first  
image of a  
linked article!



# Technical details

- Written on **Python** and **PyQt4**
- Current version: 0.9.7 (**beta**)
- size: < 1500 loc
- uses Nodebox library (for Wikipedia parsing)
- **Packaging:** **py2exe** for Windows (bundles python plus all libraries used by indywiki (pyqt, re, xml.dom etc) in one executable. results in ~30 mb size (10compressed)! But it works!





# Technical details

- ...
- `from PyQt4 import QtCore, QtGui` (*PyQt4 powers the gui*)
- `import urllib,urllib2` (*fetch wikipedia data*)
- `import threading` (*python threading, for asynchronous images display etc*)
- `from web1 import wikipedia` (*Nodebox library is used to get most data*)
- `from xml.dom import minidom` (*parses the xml data, that Mediawiki api gives us*)
- ...



# The GUI

- Several classes and functions
- **Asynchronous** display of images (achieved through the use of python **threads**)
- Many emits, because of the xlib first rule (“**You do not make X calls from more than one thread**” - only from the main)



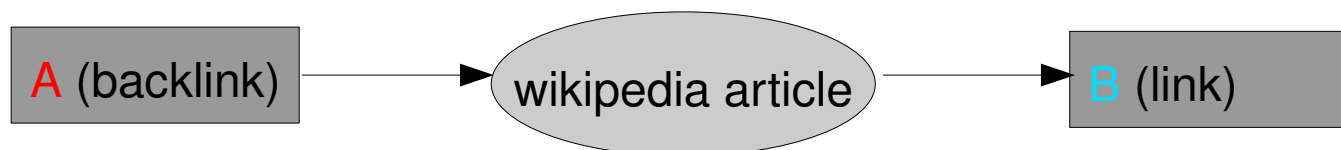
# class Ui\_MainWindow

Initiates all widgets, which are:

- centralwidget,
- 10 pushbuttons (**thumbnails**),
- misc include lineedits, pushbuttons, textbrowsers
- Scheduled **toolbar** with icons (next, back etc) to substitute all buttons with icons



# Links filtering/scoring



scores:

A: 2,  
B:2,  
{C,D...X}:1 ,  
A+B:4,  
A+B+C:5,  
A+D:3

C (belongs to category1)  
D (belongs to category2)  
.  
.  
X (belongs to category xx)

*WE GET:*

*A, {C,D...X}-->Wikipedia api  
B-->through Nodebox*

the higher the score of a link, the bigger position on the dictionary (will be displayed earlier)

***Note:The algorithm can be improved!***



# Icons or buttons?

Indywiki: exploring wikipedia visually

File View Help

CONTENTS

[Zope](#)  
[History](#)  
[Technical features](#)  
[Zope 3](#)  
[Five and the Future of Zope](#)  
[Zope Page Templates](#)  
[References](#)

Zope 3 is based on a [component architecture](#) that makes it easy to mix software components of various origins written in [Python](#). The first production release of the new software, Zope X3 3.0.0, was released on [November 6, 2004](#). The current production release, Zope 3.3.1, was released on [January 14, 2007](#).

### Five and the Future of Zope

The [Five project](#) is a Zope 2 extension that allows usage of Zope 3 technologies within Zope 2.

Originally released separately, Five has become an integral part of the main line of Zope 2 development. Five 1.0 was integrated with Zope 2 starting with Zope 2.8.0. Each successive Zope 2 release has integrated more of the features of Zope 3.

The name Five is a bit of a joke: Zope 2 + Zope 3 = Five.

In one [likely future scenario, proposed by Jim Fulton](#), Zope 3 may re-focus on providing multiple independent infrastructure components, while Zope 2 and Five would become one project focused on providing a web application server framework (an ancillary goal is to reduce confusion about what exactly "Zope" is).

### Zope Page Templates

As mentioned previously, Zope Page Templates are themselves XHTML documents which means they can be

LINKS

[2004](#)  
[2007](#)  
[Content management](#)  
[ERPS](#)  
[Fredericksburg, Virginia](#)  
[Guido van Rossum](#)  
[HTML](#)  
[Hadar Pedhazur](#)  
[Information hiding](#)  
[Internationalization and localization](#)  
[January 14](#)  
[METAL](#)  
[November 6](#)  
[Nuxeo CPS](#)  
[Plone \(content management system\)](#)  
[Python \(programming\)](#)

Indywiki search on <http://en.wikipedia.org>. Article Zope has 1 images.

90%

(problem is  
this  
distracts  
the eye)

# Custom widgets:Thumbnail buttons



```
appwidth = int(QtGui.QApplication.desktop().availableGeometry().width())
```

```
class MYB(QtGui.QPushButton):
```

```
...
```

```
    def sizeHint(self):
```

```
        if int(appwidth) == 800:
```

```
            return QtCore.QSize(140,110)
```

```
#while initiating the app:
```

```
if int(appwidth) == 800: self.pushButton_5.setIconSize(QtCore.QSize(130,100))
```



# ...Continued

```
QtCore.QObject.connect(self.pushButton_5,QtCore.SIGNAL("clicked()"),  
self.button5_clicked,QtCore.Qt.QueuedConnection) \
```

```
def button5_clicked(self):  
    if type(self.dicta['5']) is list:  
        self.display_native_size(...)  
    elif type(self.dicta['5']) is tuple:  
        self.lineEdit.clear()  
        self.lineEdit.insert(QtCore.QString(self.dicta['5'][0]))  
        self.centralwidget.first_click()
```



# Custom widgets: Text Browser

- Displays article text
- Subclass of QTextBrowser class, so that text titles of a wikipedia page link to their paragraph text

#displayed on textedit2

```
class MyTextBrowser(QtGui.QTextBrowser):
```

```
...
```

```
    def setSource(self, number_of_title):
```

```
        ...
```

```
        ui.textEdit_2.insertPlainText(ui.aquery.paragraphs[int(number_of_title.toString())])
```





# ...Continued

- links found on a wikipedia page become our search keywords

```
class MyTextBrowser2(QtGui.QTextBrowser):
```

```
...
```

```
    def setSource(self, title):
```

```
        ...
```

```
        ui.lineEdit.insert(QtCore.QString(title.toString()))
```

```
        ui.centralwidget.first_click()
```



# Nice things to do...

- 'Port' the application for the **web**
- display **HTML** (**not as easy as it sounds**)
- improve overall **aesthetics**
- add caching capabilities
- get responses from **multiple** wikipedia sites
- add **slide show** capabilities
- show the **relationship** between the images (academic)

And for the end...

make the Wikipedia images (similar to google images)!

# More on the last bullet



Wikiwix - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.wikiwix.com/index.php?img=true&action=athens&boolop=and&lang=en&disp=article

Getting Started FS daily RKWard W Wikipedia:Tools - Wi... W Wikipedia:Researchi... W Wikipedia:Wikipedia I... W Wikipedia:Researchi... SOFTWARE GREECE

Gmail: Email from Google groovesalad.pls (audio/... Wikiwix

wikiwix the ultimate Wikipedia articles search engine

Articles Images Atlas

athens

Search with wikiwix

with all of the words with at least one of the words

I love WIKIPEDIA

Results 1 - 20 of 50 for athens 2244181 articles , 10764 desynchronized articles

Acropolis from south west

Athena type Velletri

EarlyAthenianCoin

Mount Lycabettus

AtheneOudheid

Ac. academy

DionysiusTheater

Epidaurus Theater

TragicComicMasksHadriansVillamosaic

Parthenon from south

Leonidas statue1b

Acropolis of Athens 01361

Ac. templezeus

Stoa in Athens

Athens Acropolis

Model temple of Aphaia Glyptothek Munich

Done

Tor Disabled

Now: Mostly Cloudy, 7° C

Tue: 8° C

Wed: 7° C

“ Something similar to what wikiwix is doing, but use indywiki's algorithm to score links.”

# Community



- Such a nice thing!
- a few thousands downloads
- people mention indywiki on blogs, articles and send suggestions
- indywiki appears on kde-apps.org, sourceforge and freshmeat
- Still no code contribution (that's why I'm at Fosdem) :)

Thank you for your time!



**Questions and Demo**

<http://indywiki.sf.net>