

2. ALL ROADS LEAD TO ROME

Team Croatia

Reporter: Andrej Todic



2. ALL ROADS LEAD TO ROME

Open a **random** Wikipedia article and click on the **first** link in the article. Keep clicking on the **first** link of each following article. It is argued that you will quickly end up on the page **Philosophy**. **Investigate** whether this is true. How can one **describe** such an observation?

OUTLINE



Theoretical introduction

- Facts about Wikipedia
- Special: Random Feature



Experiment

- Use of Programming
- Loop Occurrences



Results

- Results (with and without modifications)
- Description

WIKIPEDIA

- Multilingual web-based encyclopedia
- Articles can be modified by volunteers
- 261 different languages
- 15 million pages → about 5 million in English



WIKIPEDIA
The Free Encyclopedia

EXAMPLE

Science

From Wikipedia, the free encyclopedia

This article is about the general term. For other uses, see [Science \(disambiguation\)](#).

Science (from **Latin** *scientia*, meaning "knowledge")^{[2][3]:58} is a systematic enterprise that builds and organizes knowledge in the form of testable [explanations](#) and [predictions](#) about the [universe](#).^[a]



Latin

From Wikipedia, the free encyclopedia
(Redirected from [Latin language](#))

For other uses, see [Latin \(disambiguation\)](#).

Latin (Latin: *lingua latīna*, IPA: [ˈlɪŋɡw̥a laˈtiːna]) is a [classical language](#) belonging to the [Italic branch](#) of the [Indo-European languages](#). The [Latin alphabet](#) is derived from the [Etruscan](#) and [Greek alphabets](#), and ultimately from the [Phoenician alphabet](#).

SPECIAL:RANDOM

- Generates random articles
- Each page in the Wikipedia database has a unique index
- An article is chosen by choosing a random index
- Used to take random samples from the Wikipedia database

/wiki/Special:Random



random Wikipedia article

HYPOTHESES

- **Hypothesis 1:**

More than half of all articles will lead to *Philosophy*

- **Hypothesis 2:**

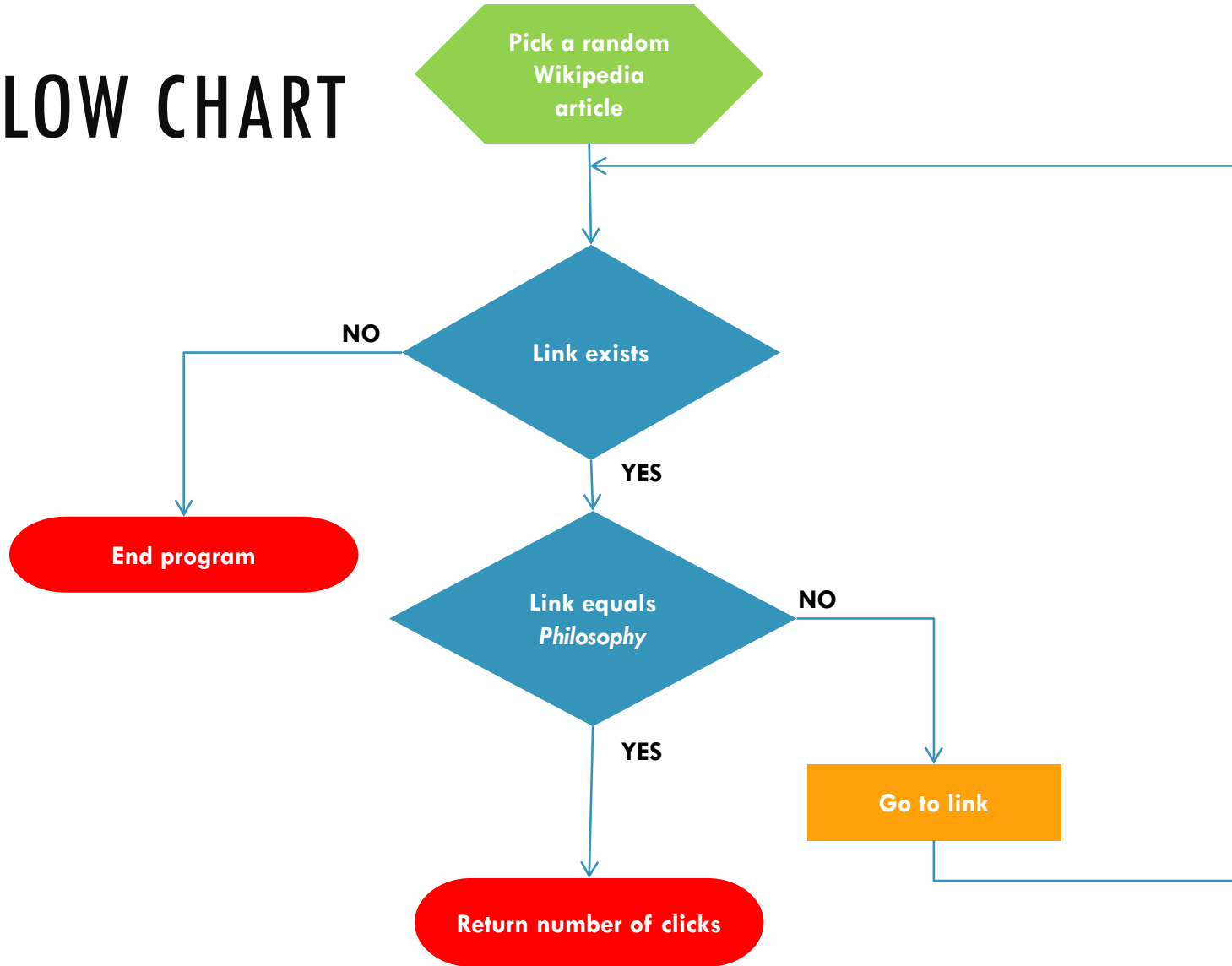
If it's possible to reach *Philosophy*, it can be done in **under 50 clicks**

TESTING

- Automated using **Programming**
- Web **client** using the programming language Python
- **Automatically** searches Wikipedia



FLOW CHART



TESTING

- Program checks if randomly picked web pages lead to *Philosophy*
- Output consists of
 - Address of the starting web page
 - Number of clicks to reach *Philosophy*
- Wikipedia article doesn't lead to Philosophy

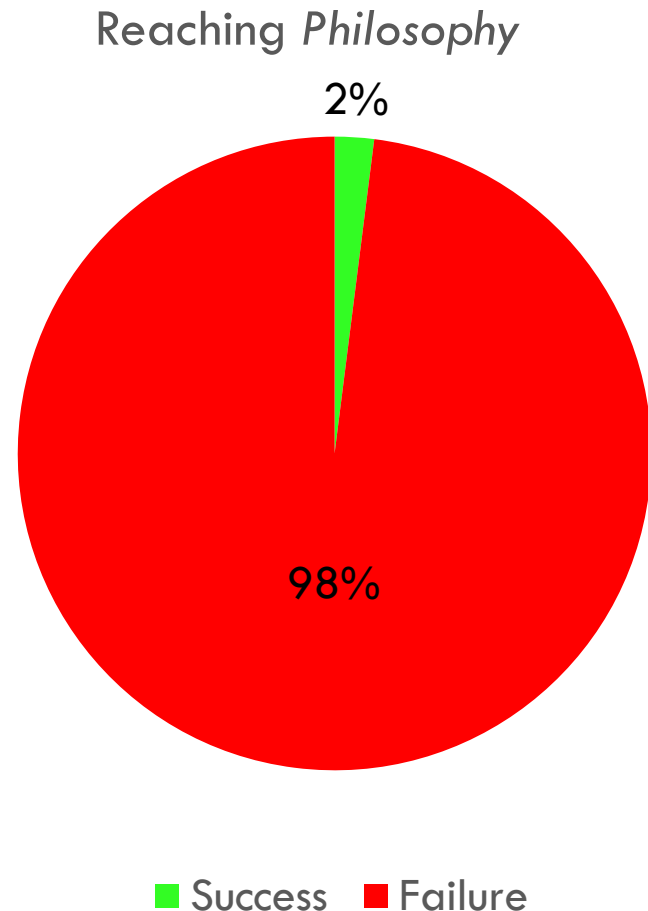


Number of clicks is set to -1

INFINITE LOOPS

First links of randomly chosen articles:

- lead to *Philosophy*
 - OR
 - lead to already visited articles
- Second outcome means all the articles would have to be visited again
 - Test is stuck in a loop



WHEN LOOPS OCCUR

Science

From Wikipedia, the free encyclopedia

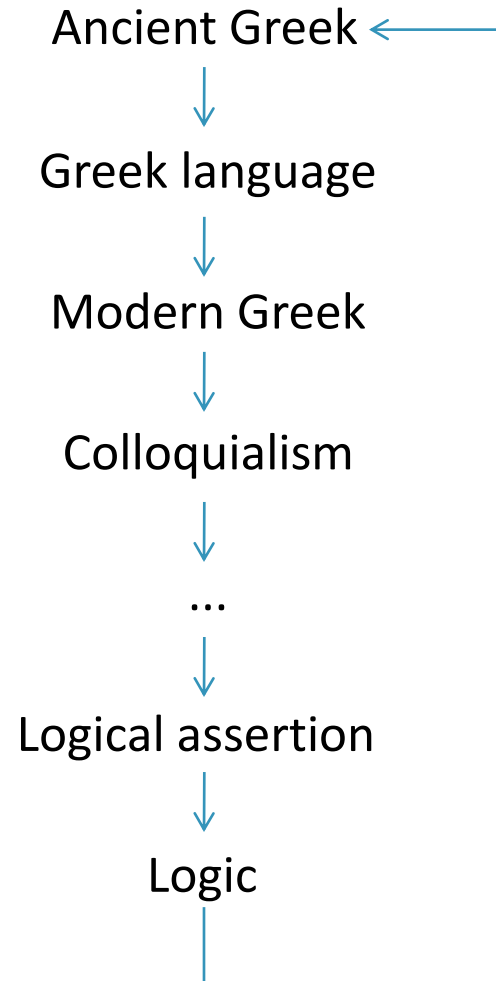
This article is about the general term. For other u.

Science (from Latin *scientia*, meaning "knowledge")^[c] testable explanations and predictions about the unive

Contemporary science is typically subdivided into the and societies, and the formal sciences like mathematical disciplines cannot be tested with physical observation engineering and medicine may also be considered to university, a college, or a research institute.

From classical antiquity through the 19th century, sci fact, in the West the term "natural philosophy" encom astronomy, medicine, among many others [7]:3[b] In th

- First link is often about the etymology of a word

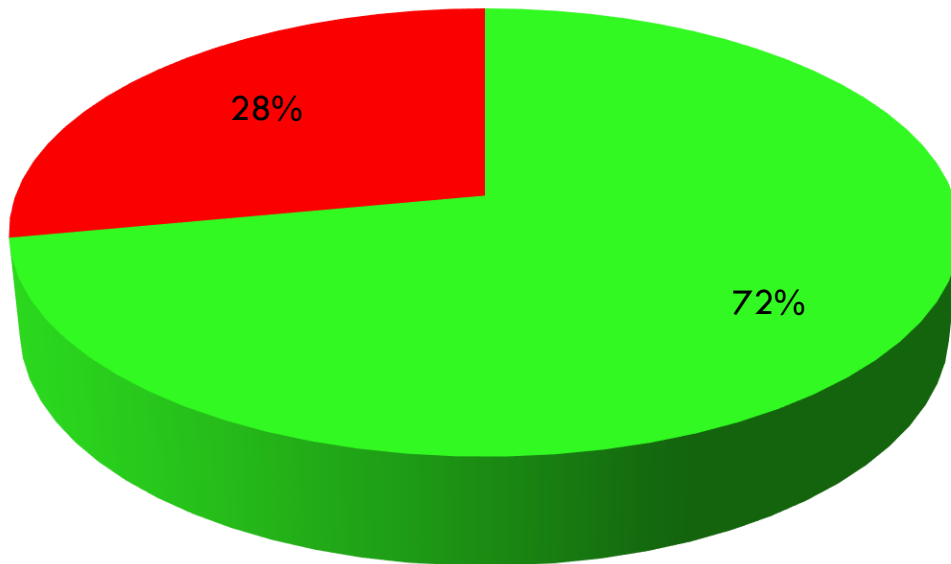


MODIFICATION

- Ignoring the first links connected to infinite loops, such as:
 - “/wiki/Mathematics”
 - “/wiki/Truth”
 - “/wiki/Greek”
 - “/wiki/Latin”
- Moving on to the next link
- Possible to reach *Philosophy* more often

RESULTS

Reaching Philosophy

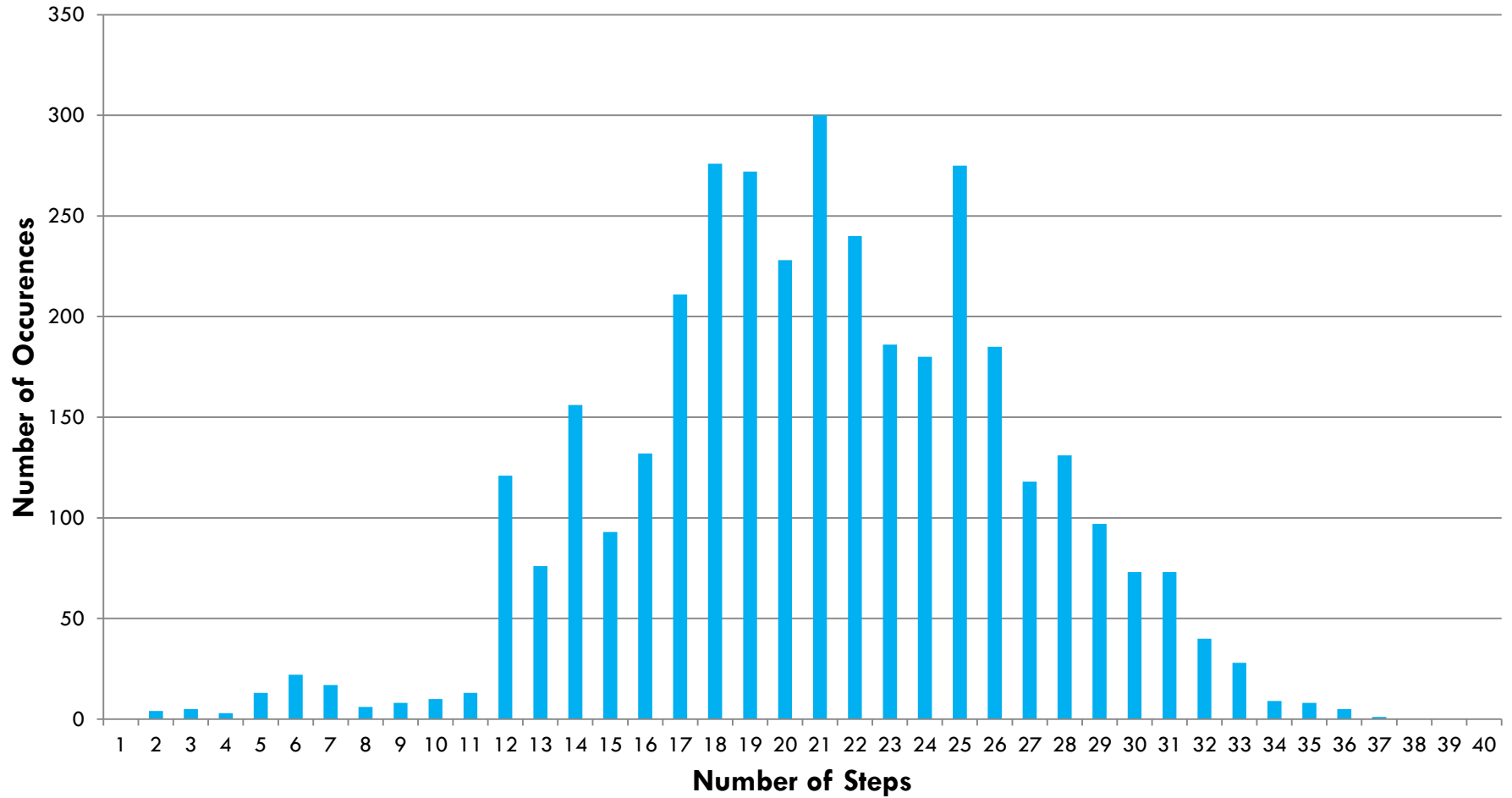


■ Success ■ Failure

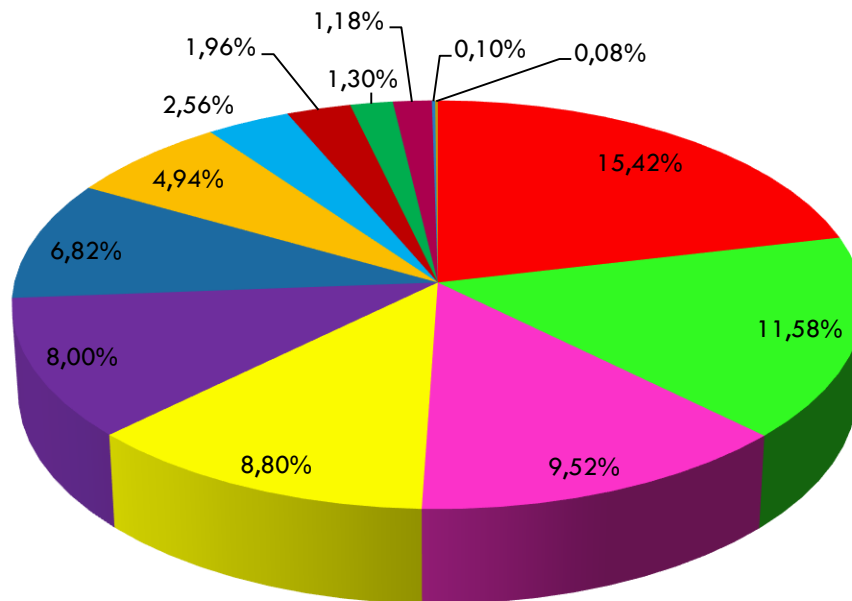
28% which fail lead to:

- Main Page
- Page without links

RESULTS



RESULTS



- Emergent properties
- A priori and a posteriori
- Property(philosophy)
- Idea
- Philosopher
- Aesthetic
- Poiesis
- Quality(philosophy)
- Ethic
- Epistemology of science
- Absurdism
- Aesthetics
- Logical form

RESULTS

- Experiment repeated for 5000 randomly chosen articles
- Average number of clicks ≈ 21

HOW CAN ONE DESCRIBE SUCH AN OBSERVATION?

Philosophy (from Greek φιλοσοφία, philosophia, literally „love of wisdom”) is the study of **general** and **fundamental** problems concerning matters such as existence, knowledge, values, reason, mind, and language.

Wikipedia

CONCLUSION

- If the task from the assignment is followed literally, it rarely leads to *Philosophy*
 - The number of links which lead to *Philosophy* is too small to make a conclusion
- If first links which lead to loops are ignored
 - Both hypotheses are correct
 - Hypothesis 1:
More than half of all articles will lead to *Philosophy*
 - Hypothesis 2:
If it's possible to reach *Philosophy*, it can be done in under 50 clicks
- Philosophy – general and fundamental

POSSIBILITIES FOR FUTURE WORK

- Downloading all Wikipedia articles for individual testing
- Drawing a tree diagram which represents all the connections to *Philosophy*
- Speeding up the program
- Testing for articles other than *Philosophy*

LITERATURE AND SOURCES

- <https://en.wikipedia.org/wiki/Wikipedia:FAQ/Technical#random>, accessed on: 01.02.2018., 17:56
- <https://computer.howstuffworks.com/internet/basics/wiki1.htm>, accessed on: 01.02.2018, 16:34
- <http://www.findingdulcinea.com/guides/Technology/Wikipedia.xa1.html>, accessed on: 01.02.2018. 15:22

THANK YOU!

Team Croatia

Reporter: Andrej Todic



ADDITIONAL SLIDES

```
import http.client
from html.parser import HTMLParser

dic = {}
link_test = ''
dic_num = {}

failure = 0
sum_of_steps = 0

file_1 = open("output_5.txt", "w")
file_2 = open("direct_links_5.txt", "w")

class MyHTMLParser(HTMLParser):
    href_var = ''
    flag_1 = 0
    num = 0
    l = []
    flag_2 = 1
```

ADDITIONAL SLIDES

```
def handle_starttag(self, tag, attrs):
    if tag == 'p':
        self.flag_1 = 1

    if tag == 'td':
        self.flag_2 = 0

    if tag == 'a' and self.flag_1 and self.flag_2:
        for name, value in attrs:
            if name == "href" and self.num == 0:
                if "IPA" not in value and "." not in value and "#" not in value:
                    if "Latin" not in value and "Greek" not in value:
                        if "Mathematic" not in value and value != "/wiki/Truth":
                            self.href_var = value
                            self.l.append(self.href_var)
                            self.num += 1

def handle_endtag(self, tag):
    if tag == 'p':
        self.flag_1 = 0

    if tag == 'td':
        self.flag_2 = 1
```


ADDITIONAL SLIDES

```
for number_of_iterations in range(10000):
    link = '/wiki/Special:Random'
    counter = 0
    flag = 0
    array = []

    while link != '/wiki/Philosophy':
        conn = http.client.HTTPSConnection("en.wikipedia.org")
        allow_redirects = True
        conn.request("GET", link)
        r1 = conn.getresponse()

        while r1.status != 200: #status 200 means that the link has been found
            link = r1.getheader('Location')
            conn.close()
            conn = http.client.HTTPSConnection("en.wikipedia.org")
            conn.request("GET", link)
            r1 = conn.getresponse()
```

ADDITIONAL SLIDES

```
if flag == 0:
    starting_link = link
    flag = 1

if link in dic_num.keys():
    counter += dic_num[link]
    break

if link in dic.keys():
    link = dic[link]

else:
    code = str(r1.read())
    parser = MyHTMLParser()
    parser.feed(code)

    link_test = parser.href_var

    if link_test == '/wiki/Philosophy':
        direct_link = link

    dic.update({link : link_test})
    link = link_test

counter += 1
conn.close()
```

ADDITIONAL SLIDES

```
    if link in array:
        counter = -1
        break

    array.append(link)

dic_num.update({starting_link : counter})
file_1.write('Address: {}'.format(starting_link))
file_1.write('\t')
file_1.write('Number of clicks: {}'.format(counter))
file_1.write('\n')

if counter == -1:
    failure += 1

else:
    sum_of_steps += counter
    file_2.write(direct_link)
    file_2.write('\n')
```

ADDITIONAL SLIDES

```
file_1.write('\n')
file_1.write('\n')
file_1.write('Number of iterations: {}'.format(number_of_iterations))
file_1.write('\n')
file_1.write('Average number of clicks: {}'.format(sum_of_steps / (number_of_iterations - failure)))
file_1.write('\n')
file_1.write('Probability of success: {}'.format((number_of_iterations - failure) / number_of_iterations))
file_1.close()
file_2.close()
```