... vs. using the library

## hw10pr3

files and the dictionary class

files

dictionaries
... vs. using the library
hw10pr3
files and the dictionary class


## If an Algorithm Wrote This, How Would You Even Know?

## Algorithmic Authorship...?

## a.txt - /Users/zdodds/Desktop/a.txt

I like poptarts and 42 and spam. Will I get spam and poptarts for the holidays? I like spam poptarts!
suppose this text represents my "style" ...

## How could a program author new prose in this same style?!

## Algorithmic Authorship...!

$\Theta \bigcirc 0$ a.txt $-/$ sers/zdodds/Desktop/a.txt
I like poptarts and 42 and spam.
Will I get spam and poptarts for
the holidays? I like spam poptarts!
suppose this text represents my "style" ...

What would be a reasonable first word to start a newlygenerated sentence?

What would be a reasonable next word to follow the first?

What would be a reasonable test for sentence-ending?

## Algorithmic authoring examples...

'Cause somethin' like he left knee and a harp," said he had to the whole school? The shouting and then some strange and Mrs. "Well, I know Hagrid; they spotted handkerchief and get him get rid of course, had a gigantic beet with her," he knew what to all he's

Wanna live while we're cool, so tonight What a feeling to be doing what I wish I know we only met but it ain't hard to be nothing left The story of my life I'm watching her eyes smile you flip your eyes You don't know what makes you got stars, they're in the wire She said, "Can I got a feeling to be a dentist

Who's the original human author of each of these?

This is but ourselves. No, faith, My uncle! O royal bed of confession Of your rue for leave to nature; to this time I should weep for thy life is rotten before he is. have sworn 't. Or my blood. I have closely sent for nine; and unprofitable,

The Senators and the date of a written declaration that Purpose, they shall consist of nine States, shall not, when he shall have such Vacancies. The President pro tempore, in the Desire of a Qualification to the Speaker of the Senate. Article 6. When vacancies by the office upon probable

## Markov Models

## Techniques for modeling any sequence of natural data

speech, text, sensor data...

1st-order Markov Model
(defining property)

## Each item depends only on the one immediately before it.

## Lists are sequential containers:

$$
I=[47, ~ 4, ~ 47, ~ 42]
$$

index

Dictionaries are arhitmWe need a new data structure! a structure! $(A$ new $\underset{\text { mon }}{\text { class.... }}$ )

## Lists are sequential containers:

$$
\pm=\left[\begin{array}{cccc}
47, & 47, & 42
\end{array}\right]
$$

## Dictionaries are arbitrary containers:

$\mathrm{d}=\{47$
key

42 :
1 \}



## Dictionaries are arbitrary containers:

zd = \{'rabbit':1999,
value
'ox':1997\}
key

elements (or values) are looked up by a key starting anywhere you want! Keys don't have to be ints!

What's zd's data here?

Now I see the key to dictionaries' value...

## Dictionaries are arbitrary containers:


elements (or values) are looked up by a key starting anywhere you want! Keys don't have to be ints!

| Rat | Feb 19 1996-Feb 06 1997 |
| :--- | :--- |
| Ox | Feb 07 1997-Jan 271998 |
| Tiger | Jan 28 1998-Feb 151999 |
| Rabbit | Feb 16 1999-Feb 042000 |
| Dragon | Feb 05 2000-Jan 232001 |
| Snake | Jan 24 2001-Feb 112002 |
| Horse | Feb 12 2002-Jan 312003 |
| Goat | Feb 01 2003-Jan 212004 |
| Monkey | Jan 22 2004-Feb 082005 |
| Rooster | Feb 09 2005-Jan 282006 |

12-year zodiac...
Now I see the key to dictionaries' value...

## Dictionaries are arbitrary containers:

$z=\{' r a b b i t ':[1999,1987,1975, \ldots]$,

> 'ox': [1997,1985,1973,...], 'tiger': [1998,2010,...], ... \}

What type are the keys?

What type are the values?

## Dictionaries are arbitrary containers:

$z=\{' r a b b i t ':[1999,1987,1975, \ldots]$,
'ox': [1997,1985,1973,...],
'dragon': [2000,1988,1976,...],

$$
\ldots\}
$$

Is 'dragon' a key in $\mathbf{z}$ ?
if 'dragon' in $\mathbf{z}$

Is 1969 in
z['dragon']?

```
LoW = [ 'spam', 'spam', 'poptarts', 'spam' ]
                                w will be...
    d[w] += 1
d will be...
    {}
```

d = {}

```
d = {}
    if w not in d:
    if w not in d:
    d[w] = 1
    d[w] = 1
    else:
```

    else:
    ```
```

    for w in LoW:
    ```
```

    for w in LoW:
    ```
```

    for w in LoW:
    ``` w will be...
\{'poptarts':1, 'spam':3\}
```

LoW = [ 'spam', 'spam', 'poptarts', 'spam' ]
d={} w will be...
for w in LoW:
if w not in d:
d[w] = 1
else:
d[w] += 1

```

\section*{d will be...}
\{\}
```

$$
\begin{array}{ll}
w=\text { 'spam' } & \{\text { 'spam':1\} } \\
w=\text { ='spam' } & \{\text { 'spam':2\} } \\
w=\text { ='poptarts'' } & \{\text { 'poptarts':1, 'spam':2\} } \\
w=\text { ='spam' } & \{\text { 'poptarts':1, 'spam':3\} }
\end{array}
$$

final d

```

\section*{LoW = [ 'spam', 'spam', 'poptarts', 'spam' ]}
but where to get d will so many words?
d = {}
d = {}
for w in LoW:
for w in LoW:
    if \(w\) not in \(d\) :
        \(d[w]=1\)
    else:
    d[w] += 1

\section*{Files}

In Python reading files is smooth...

text \(=\mathrm{f}\). read ()
reads the whole file into the string text

\section*{f.close()}
closes the file (optional)

\section*{text}
'I like poptarts and 42 and spam. V nWill I

LoW = text.split()
[ 'I', 'like', 'poptarts', ... ]

\section*{def word_count( filename ):}
```

f = open( filename )
text = f.read()
f.close()
file handling
LoW = text.split()
print("There are",len(LoW),"words")

```

What if we wanted the number of different words in the file?

This would be the author's
vocabulary count, instead of the total word count.

\section*{Vocabulary, anyone?}

Shakespeare used 31,534 different words -- and a grand total of 884,647 words, counting repetitions across all of his works....

Shakespearean coinages gust besmirch
unreal
superscript
watchdog
swagger
successful
affined rooky attasked out-villained
unsuccessful

There's also one contemporary British author in the Oxford English Dictionary...

\section*{Who?}
with what word?

\section*{Vocabulary, anyone?}

Shakespeare used 31,534 different words - - and a
grand total of 884,647 words, counting repetitions
across all of his works....

Shakespearean coinages
> gust
> besmirch
> unreal
> superscript
> watchdog

swagger
successful
affined rooky attasked out-villained
unsuccessful

\section*{muggle}

\section*{'Muggle' goes into Oxford English Dictionary}

JK Rowling's word for non-wizards "muggle" - has made it into the new edition of the Oxford English Dictionary (OED).

The draft definition according to the dictionary's website says:
- Muggle: invented by JK (Joanne Kathleen) Rowling (b. 1965), British author of children's fantasy fiction (see quot. 1997).

In the fiction of JK Rowling: a person who possesses no magical powers. Hence in allusive and extended uses: a person who lacks a particular skill or skills, or who is regarded as inferior in some way.
J. K. Rowling
```

from filename import defaultdict
def vocab_count( filename ):
f = open( filename ) }
file handling
text = f.read()
f.close()
LoW = text.split()
print "There are", len(LOW), "words."
d = {}
for w in LoW:
if w not in d:
d[w] = 1
else:
d[w] += 1
print "There are", len(d), "_distinct_ words.\n"
return d \# return d for later use by other code...

```

\section*{Markov Models can be generative!}

A key benefit of Markov Models is that they can generate feasible data!

\section*{Original file:}

I like poptarts and 42 and spam.
Will I get spam and poptarts for the holidays? I like spam poptarts!
```

d = create_model('hpwhich.txt')
d = create_model('randj.txt')
d = create_model('oneD.txt')\
d = create_model('a.txt')
gt(d, 250)

```

\section*{Markov Models can be generative!}

A key benefit of Markov Models is that they can generate feasible data!
Original file:
I like poptarts and 42 and spam.
Will I get spam and poptarts for
the holidays? I like spam poptarts!

Generated text:

I get spam poptarts! I like poptarts and 42 and spam. I like spam and 42 and 42 and 42 and spam. Will I like poptarts and 42 and poptarts and 42 and poptarts and 42 and 42 and poptarts and spam. I get spam and 42 and 42 and...

\section*{Our Markov Model}

Try it!
keys values
Original file

Markov Model
A dictionary!

What are the keys?
What are the values?

What are the missing values?

What is the '\$'?

Why do some keys seem missing?
'\$': ['I', 'will', 'r'],
'I': ['like', 'get', 'like']
'like':
'poptarts': ['and', 'for'],
'and': ['42', 'spam.', 'poptarts'],
'42': ['and'],
'Will': ['I'],
'the':
'spam': ['and', 'poptarts!'],
'get': ['spam'],
'for': ['the']


\section*{Our Markov Model}

Try it!
keys values
Original file

Markov Model
A dictionary!

What are the keys?

> What are the values?

What are the missing values?

What is the '\$'?

Why do some keys seem missing?
'\$': ['I', 'Will', 'I'],
'I': ['like', 'get', 'like']
'like': ['poptarts', 'spam'], 'poptarts': ['and', 'for'], 'and': ['42', 'spam.',' 'poptarts'], '42': ['and'], 'Will': ['I'],
'the': ['holidays?'],
'spam': ['and', 'poptarts!'], 'get': ['spam'],
'for': ['the']

dictionary's end

\section*{Markov-modeling's algorithm}

LoW ['I','like','spam.','I','eat','poptarts!']
pw
nw
```

d = {}
pw = '\$'
for nw in LoW:
if pw not in d:
d[pw] = [nw]
else:
d[pw] += [nw]
pw =

```
    like : [ spam.]
    eat : [ poptarts! ]

\section*{Model creation:}
1) start with the previous word, pw as '\$'
2) for each next word, \(n w\), in the list of words, add it in ...
3) then change pw to nw ...
(a) except if \(\mathbf{n w}[-1]\) was punctuation: change \(\mathbf{p w}\) to...

\section*{Generating text:}
1) start with pw as the '\$' string
2) choose a nw that follows pw, at random.
3) print nw, (the comma continues on the same line)
4) pw gets set to either nw or '\$' or if \(\mathbf{n w}[-1]\) was punctuation: change pw to...

\section*{Generating prose? Academic Opportunity!}

Confidential Paper Destruction - SergeantShredder.com - Professional secure document Shredding Services.


\section*{2nd CFP - Systemics, Informatics and Cybernetics Inbox |x}
WMSCI 2011 to DODDS
Dear Zachary Di
We invite you to subm. Japer/abstract to The 15th World Multi-Conference on Systemics, Oct 24 Reply
Cybernetics and Informatics: WMSCI 2011, to be held in Orlando, Florida, USA, on July 19th - July
22nd, 2011 (www.2011iiisconferences.org/wmsci)
If you have any colleagues who might be interested in making a submission to the conference,
please feel free to forward this e-mail to them.
Below are the next deadlines for WMSCI 2011 (Check the web site for possible extensions or new
set of deadlines):
Papers/Abstracts Submission and Invited Session Proposals: November 25th, 2010
Authors Notifications: January 31st, 2011
Camera-ready, full papers: February 28th, 2011

\section*{WMSCI}

\section*{CFP (deadline extension) - Robotics (July 8-11, 2017)}
! WMSCI 2017 cfp-summer@mail.iiisconf2017.org via cs.hmc.edu
1:19 AM (5 minutes ago) \(\square\) to doddsWhy is this message in Spam? It contains content that's typically used in spam messages. Learn more Dear Zachary Dodds,

We would like to inform you that we extended to * April 5, 2017 * the submission deadline for your potential con: .oution in the area "Robotics" or any other included in the \(21^{\text {st }}\) World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2017 (http://www.2017iiisconf.org/wmsci), to be held on July 8-11, 2017, in Orlando, Florida, USA, jointly with:
- The \(11^{\text {th }}\) International Multi-Conference on Society, Cybernetics, and Informatics: IMSCI 2017
- The \(15^{\text {th }}\) International Conference on Education and Information Systems, Technologies and Applications: EISTA 2017
- The \(10^{\text {th }}\) International Multi-Conference on Engineering and Technological Innovation: IMETI 2017

The respective web sites of the above events and the others being jointly organized can be found at the general CFP posted at: http://www.2017iiisconf.org/cfp-summer2017.asp

To submit your article, please click the "Authors" tab on the conference website. Submissions for face-to-face and virtual presentations are both accepted.

WMSCI and all its collocated events are being indexed by Elsevier's SCOPUS since 2005. The 2017 proceedings will also be sent to Elsevier's SCOPUS.

\section*{WMSCI}


\section*{WMSCI}


\section*{WMSCI 2005}

\section*{Rooter: A Methodology for the Typical Unification of Access Points and Redundancy}

Markov-generated submission accepted to WMSCI 2005

Not a first-order model ... but a third-order model

\title{
Rooter: A Methodology for the Typical Unification of Access Points and Redundancy
}

\author{
Jeremy Stribling, Daniel Aguayo and Maxwell Krohn
}

\begin{abstract}
Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and publicprivate key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interposable.
\end{abstract}

\section*{I. Introduction}

Many scholars would agree that, had it not been for active networks, the simulation of Lamport clocks might never have occurred. The notion that end-users synchronize with the investigation of Markov models is rarely outdated. A theoretical grand challenge in theory is the important unification of virtual machines and real-time theory. To what extent can web browsers be constructed to achieve this purpose?

Certainly, the usual methods for the emulation of Smalltalk that paved the way for the investigation of rasterization do not apply in this area. In the opinions of many, despite the fact that conventional wisdom states that this grand challenge is continuously ancwered hy the study of access noints we

The rest of this paper is organized as follows. For starters, we motivate the need for fiber-optic cables. We place our work in context with the prior work in this area. To address this obstacle, we disprove that even though the muchtauted autonomous algorithm for the construction of digital-to-analog converters by Jones [10] is NP-complete, objectoriented languages can be made signed, decentralized, and signed. Along these same lines, to accomplish this mission, we concentrate our efforts on showing that the famous ubiquitous algorithm for the exploration of robots by Sato et al. runs in \(\Omega((n+\log n))\) time [22]. In the end, we conclude.

\section*{II. Architecture}

Our research is principled. Consider the early methodology by Martin and Smith; our model is similar, but will actually overcome this grand challenge. Despite the fact that such a claim at first glance seems unexpected, it is buffetted by previous work in the field. Any significant development of secure theory will clearly require that the acclaimed realtime algorithm for the refinement of write-ahead logging by Edward Feigenbaum et al. [15] is impossible; our application ic no differant Thic mav ar mav nat actally hald in ranlity


\section*{the Typical Unification d Redundancy}

\author{
and Maxwell Krohn
}
rest of this paper is organized as follows. For starters, otivate the need for fiber-optic cables. We place our in context with the prior work in this area. To adthis obstacle, we disprove that even though the muchautonomous algorithm for the construction of digitallog converters by Jones [10] is NP-complete, objected languages can be made signed, decentralized, and d. Along these same lines, to accomplish this mission, we ntrate our efforts on showing that the famous ubiquitous thm for the exploration of robots by Sato et al. runs in \(+\log n)\) ) time [22]. In the end, we conclude.

\section*{II. ARCHITECTURE}
research is principled. Consider the early methodology artin and Smith; our model is similar, but will actually ome this grand challenge. Despite the fact that such m at first glance seems unexpected, it is buffetted by pus work in the field. Any significant development of theory will clearly require that the acclaimed realalgorithm for the refinement of write-ahead logging by rd Feigenbaum et al. [15] is impossible; our application different Thic mav ar mav not actully hald in reality

\section*{There are no one-sided coins...}

\section*{iOS Just Got A Paper On Nuclear Physics Accepted At A Scientific Conference \\ Posted by Christoph Bartneck on Oct 20, 2016 in Featured, Research | 7 comments}


Automatically generating scientific articles has become easy with dedicated software such as SCIgen. Even a paper that only repeated the sentence "Get me of your mailing list" was recently accepted for publication. Today I received an invitation from the International Conference on Atomic and Nuclear Physics to submit a paper. Since I have practically no knowledge of Nuclear Physics I resorted to iOS auto-complete function to help me writing the paper. I started a sentence with "Atomic" or "Nuclear" and then randomly hit the auto-complete suggestions. The text really does not make any sense. After adding the first illustration on nuclear physics from Wikipedia, some references and creating a fake identity (Iris Pear, aka Siri Apple) I submitted the paper which was accepted only three hours later! I know that IOS is a pretty good software, but reaching tenure has never been this close.

UPDATE (27/10/2016): Turns out that conference organizer, OMICS Group, is currently under federal investigation.


\section*{Thesis deadlines?}

\section*{Papers due?}

Have Python write your papers for you...```

