SELECT * FROM culture
ORDER BY random() LIMIT 1;

Old School Flyers @oldschoolflyers · Mar 14

Talent & Disco Birthday Bash!
Given by
The Deuce
Lil Rodney-C K.K. Rockwell
Double Trouble

“Super Birthday Party”
“Super City Wide Talent Show”

A talk by Leonard Richardson (@leonardr)
http://www.crummy.com/
I love collection selection bots!

[Studio portrait of a dog.]
digitalcollections.nypl.org/items/510d47e0 ...
They have artistic, aesthetic and educational value

Moon Shot Bot @moonshotbot · 13h
Apollo 7 Hasselblad image from film magazine 3/M - Earth Orbit
flickr.com/photos/project …
The audience stays engaged (versus generative bots)
All you need is a collection in a database

```sql
SELECT * from culture
ORDER BY random()
LIMIT 1;
```
<table>
<thead>
<tr>
<th>Name</th>
<th>Collection</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anita Donut (#0)</td>
<td>@DeathBot3000</td>
<td>Mar 15</td>
</tr>
<tr>
<td>Rita Haymaker</td>
<td>@DeathBot3000</td>
<td>Mar 15</td>
</tr>
<tr>
<td>Annie Ouchley</td>
<td>@DeathBot3000</td>
<td>Mar 14</td>
</tr>
<tr>
<td>Outta Daway</td>
<td>@DeathBot3000</td>
<td>Mar 14</td>
</tr>
</tbody>
</table>

What if the collection is not in a database?
What if you start with a bunch of books?

Fragmentation Damage vs. Large Targets: Large objects may be hit with multiple fragments. An object will be attacked once for each +1 Size Modifier (see p. B201) over 0. Thus, for an object with a Size Modifier of +4 (a linear measurement of 10 yards), four fragment attacks are made against it - roll individually for each attack.

Source: *Compendium II (GURPS)*

Your breathing becomes labored as the air seems to thin. You find yourself becoming exhausted more easily as you negotiate your way endlessly upward. Each step is an effort to avoid twisting an ankle or losing your foothold, and falling.

Source: *TSR 2416 - DSE1 - Dragon's Crown (Dark Sun)*
What if most of the collection is boring?

Minecraft Signs @MinecraftSigns · Feb 27
Sit n Just
Pull levers the
When Pig Spawns

Minecraft Signs @MinecraftSigns · Feb 27
Eat
and you shall be
INVINCIBLE

Minecraft Signs @MinecraftSigns · Feb 27
"Hey i may be the mayor but that dosent mean im rich!"
Make it bot-ready with a data pipeline!

1. Fetch the data
2. Chop it into bot-sized pieces
3. Filter the crap
4. Let it run for 20 years!
The lingua franca: NDJSON

["Graveyard Gertie", "13", "2011-03-19", "Circle City Derby Girls"]
["Ms Pummel Her", "007", "2010-06-07", "Cherry City Derby Girls"]
["Dolly Trauma", "", "", "No Coast Derby Girls"]
["Louiville Slugga", "1117", "2012-06-03", "Canberra Roller Derby"]
...

One JSON object per line
1. Write a script that turns the raw data into a NDJSON file.

2. Write a script that takes that NDJSON file as input, and writes another NDJSON file.

3. Write a script that takes *that* NDJSON file as input, and writes another NDJSON file.

    ...

N. Write a script that pops the top item off the final NDJSON file, formats it, and posts it.
Pipeline components fetch the data

- Download web pages...
  - ...and then scrape them.
- Ask the GitHub API about Git repositories...
  - ...and then clone them.
- Get the data somehow...
  - ...and put it into an NDJSON file.
  - Or at least put the metadata in there.
Pipeline components

crop the data

- Scrape useful information from an HTML file
- Split a PDF into page images
- Cut sentence fragments from a paragraph
- Turn a text file into a word-frequency histogram
- Extract text from a Minecraft world
- Parse a source code file into lines of text
- The output always includes a new NDJSON file!
Pipeline components filter the data

• Randomize the stream to avoid bias in a later component

• Remove items that
  - Are too big or too small
  - Are obviously offensive
  - Are duplicates, near-duplicates, or repetitive
  - Don't fit the theme of the bot

• The output always includes a new NDJSON file!
Advantages of the pipeline approach

- Leaves a record of each step
- Parallelizable
- Extensible
- Saves time when you change something
- Prevents duplicate posts
Bot credits

- @oldschoolflyers (by Darius Kazemi)
- @nypl_dogs (by Ashley Blewer)
- @moonshotbot (by Darius Kazemi)
- @MOMA_robot (by John Emerson)
- @DeathBot3000
- https://lonelydungeon.tumblr.com/
- @MinecraftSigns

More collection selection bots

- @ConwaysLives
- @pony_strategies
- @MythologyBot (by @BooDooPerson)
- @oblique_questions (by Matthew McVickar)
- @worldsofzzzt (by @doscooon)
SELECT * FROM culture
ORDER BY random() LIMIT 1;

A talk by Leonard Richardson (@leonardr)
http://www.crummy.com/
I love collection selection bots!

I love collection selection bots!

NYPL Dogs
@nypl_dogs

[Studio portrait of a dog.]
digitalcollections.nypl.org/items/510d47e0 ...
They have artistic, aesthetic and educational value.
The audience stays engaged (versus generative bots)
All you need is a collection in a database

SELECT * from culture
ORDER BY random()
LIMIT 1;
What if the collection is not in a database?
What if you start with a bunch of books?

Fragmentation Damage vs. Large Targets: Large objects may be hit with multiple fragments. An object will be attacked once for each +1 Size Modifier (see p. B201) over 0. Thus, for an object with a Size Modifier of +4 (a linear measurement of 10 yards), four fragment attacks are made against it - roll individually for each attack.

Source: Compendium II (GURPS)

Your breathing becomes labored as the air seems to thin. You find yourself becoming exhausted more easily as you negotiate your way endlessly upward. Each step is an effort to avoid twisting an ankle or losing your foothold, and falling.

Source: TSR 2416 - DSE1 - Dragon's Crown (Dark Sun)
What if most of the collection is boring?

Minecraft Signs @MinecraftSigns - Feb 27
Sit n Just
Pull levers the
When Pig Spawns
↩️ 3 3 3

Minecraft Signs @MinecraftSigns - Feb 27
Eat
and you shall be
INVINCIBLE
↩️ 4 3

Minecraft Signs @MinecraftSigns - Feb 27
"Hey i may be the mayor but that dont mean im rich!"
Make it bot-ready with a data pipeline!

1. Fetch the data
2. Chop it into bot-sized pieces
3. Filter the crap
4. Let it run for 20 years!
The lingua franca: NDJSON

["Graveyard Gertie", "13", "2011-03-19", "Circle City Derby Girls"]
["Ms Pummel Her", "007", "2010-06-07", "Cherry City Derby Girls"]
["Dolly Trauma", "", "", "No Coast Derby Girls"]
["Louiville Slugga", "1117", "2012-06-03", "Canberra Roller Derby"]
...

One JSON object per line
My process

1. Write a script that turns the raw data into a NDJSON file.
2. Write a script that takes that NDJSON file as input, and writes another NDJSON file.
3. Write a script that takes \textit{that} NDJSON file as input, and writes \textit{another} NDJSON file.
   ...
N. Write a script that pops the top item off the final NDJSON file, formats it, and posts it.
Pipeline components fetch the data

• Download web pages...
  – ...and then scrape them.

• Ask the GitHub API about Git repositories...
  – ...and then clone them.

• Get the data somehow...
  – ...and put it into an NDJSON file.
  – Or at least put the *metadata* in there.
Pipeline components
crop the data

- Scrape useful information from an HTML file
- Split a PDF into page images
- Cut sentence fragments from a paragraph
- Turn a text file into a word-frequency histogram
- Extract text from a Minecraft world
- Parse a source code file into lines of text
- The output always includes a new NDJSON file!
Pipeline components
filter the data

• Randomize the stream to avoid bias in a later component

• Remove items that
  – Are too big or too small
  – Are obviously offensive
  – Are duplicates, near-duplicates, or repetitive
  – Don't fit the theme of the bot

• The output always includes a new NDJSON file!
Advantages of the pipeline approach

- Leaves a record of each step
- Parallelizable
- Extensible
- Saves time when you change something
- Prevents duplicate posts
Bot credits

- @oldschoolflyers (by Darius Kazemi)
- @nypl_dogs (by Ashley Blewer)
- @moonshotbot (by Darius Kazemi)
- @MOMA_robot (by John Emerson)
- @DeathBot3000
- https://lonelydungeon.tumblr.com/
- @MinecraftSigns

More collection selection bots

- @ConwaysLives
- @pony_strategies
- @MythologyBot (by @BooDooPerson)
- @oblique_questions (by Matthew McVickar)
- @worldsofzzt (by @doscoon)