chatterbot-corpus Documentation

Release 1.0.0

Gunther Cox

Aug 24, 2020

Contents:

The ChatterBot Corpus is a project containing user-contributed dialog data that can be used to train chat bots to communicate.

CHAPTER 1

Data Format

The data file contained in ChatterBot Corpus is formatted using YAML syntax. This format is used because it is easily readable by both humans and machines.

Table 1: Corpus Properties

Property	Required	Description
categories	Required	A list of categories that describe the conversations.
conversations	Optional	A list of conversations. Each conversation is denoted as a list.

Here is an example of the corpus data:

```yaml
categories:
- english
- greetings
conversations:
- - Hello
  - Hi
- - Hello
  - Hi, how are you?
  - I am doing well.
- - Good day to you sir!
  - Why thank you.
- - Hi, How is it going?
  - It's going good, your self?
  - Mighty fine, thank you.
```

The values in this example have the following relationships.

Table 2: Evaluated statement relationships

Statement	Response
Hello	Hi
Hello	Hi, how are you?
Hi, how are you?	I am doing well.
Good day to you sir!	Why thank you.
Hi, How is it going?	It's going good, your self?
It's going good, your self?	Mighty fine, thank you.

CHAPTER 2

Using the ChatterBot Corpus with ChatterBot

If you are looking for information on how to use the `chatterbot-corpus` module with your chat bot build with ChatterBot, then you will want to take a look at the ChatterBot Documentation

CHAPTER 3

Indices and tables

- genindex
- modindex
- search