Children often ask their caregivers to find episodes of their favorite TV shows based only on a very short (and loosely relevant!) description of it ("the one where Arthur has a wiggly tooth") but video services like Netflix and Amazon don't currently provide such content-based search. Given summaries of each episode, can we use sequence embeddings to solve this retrieval problem?

Before beginning this homework, install the following libraries:

```
conda install -c huggingface transformers
pip install -U sentence-transformers
conda install -c conda-forge ipywidgets
```

First, let's read in our data for the TV show "Wild Kratts" (from Wikipedia), which has the following (tab-separated) form:

Episode	Title	Summary
1	Mom of a Croc	At the Nile River, zoologists Chris and Martin Kratt (voiced by their real-world selves) are on a mission to show one of their fellow Wild Kratts team members—brilliant young inventor Aviva Corcovado (Athena Karkanis)—that there's more to crocodiles than just violence and snapping jaws. After shrinking themselves down to a few inches tall by using Aviva's Miniaturizer invention, the Kratt Brothers disguise themselves as crocodile eggs and sneak into a mother crocodile's new nest. In the Wild Kratts team's turtle-shaped aircraft and headquarters—the Tortuga, one of Aviva's greatest inventions—the Wild Kratts tech team, consisting of Aviva, communications expert and mechanic Koki (Heather Bambrick), and skilled pilot Jimmy Z (Jonathan Malen) monitor Chris and Martin and watch as the mother crocodile faithfully guards her nest against predators for months without even eating anything. Eventually, as the crocodile eggs hatch and the crocodile mom uses her mouth to carry several of her newly hatched babies to the river, Aviva changes her mind about crocodiles and decides that these reptiles are in fact caring and dedicated mothers. But when the mother crocodile leaves the river to go get more hatchlings from her nest, predators threaten the first batch of baby crocodiles. The Kratt Brothers must use the incredible Creature Power Suits—two of Aviva's inventions—to gain the abilities of crocodiles and protect the vulnerable crocodile hatchlings.
2	Whale of a Squid	The Kratt Brothers use Aviva's amphipod-inspired submersible, the Amphisub, to dive into the deep waters of the Southern Ocean. There, they witness a never-before-seen wildlife moment: a battle between a sperm whale and a giant squid. However, the water pressure at the extreme depths where the battle is taking place badly damages and partially crushes the Amphisub, forcing Aviva to use her new ExtendoArm invention to pull the submersible back to the Tortuga. To allow Chris and Martin to return to the site of the whale-versus-squid battle, Aviva programs two new Creature Power Suits—Sperm Whale Power for Chris, and Squid Power for Martin. The Kratt Brothers use their new Creature Powers to dive back into the deep sea, where the sperm whale and the giant squid are still locked in combat. Suddenly, the sperm whale becomes entangled in a discarded fishing net and begins sinking toward an area full of underwater volcanoes. To make matters worse, a colossal squid attacks the sperm whale's calf. Chris and Martin must put their Creature Powers of both sperm whale and squid to good use to rescue the mother sperm whale and her calf.

In [1]:

```
def read_data(filename):
    data=[]
    with open(filename, encoding="utf-8") as file:
        for line in file:
            cols=line.rstrip().split("\t")
            episode=cols[0]
            title=cols[1]
            summary=cols[2]
            data.append((episode, title, summary))
    return data
```

In [2]:
```
data=read_data("../data/wild_kratts_episodes.txt")
```

In [3]:
```
def get_document_reps_for_data(data, sequence_embedding_function, model):

    # This function applies the sequence_embedding_function argument (a function itself
    # element in the input data list, and returns a copy of that list with an embedding

    data_with_reps=[]

    for episode, title, summary in data:
        data_with_reps.append((episode, title, summary, sequence_embedding_function(mod

    return data_with_reps
```

In [4]:
```
def cosine_similarity(a, b):
    return np.dot(a, b)/(np.linalg.norm(a)*np.linalg.norm(b))
```

First, we may be tempted to use the [CLS] token for BERT to represent an entire input string (as is often done in *supervised* document classification models). How well does this work as an out-of-the-box document representation not optimized for our particular task?

In [5]:
```
from transformers import BertModel, BertTokenizer
import numpy as np
from sentence_transformers import SentenceTransformer
```

In [6]:
```
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
model = BertModel.from_pretrained('bert-base-uncased')
```

```
Some weights of the model checkpoint at bert-base-uncased were not used when initializin
g BertModel: ['cls.predictions.bias', 'cls.predictions.transform.dense.bias', 'cls.predi
ctions.transform.dense.weight', 'cls.predictions.transform.LayerNorm.bias', 'cls.seq_rel
ationship.bias', 'cls.seq_relationship.weight', 'cls.predictions.transform.LayerNorm.wei
ght', 'cls.predictions.decoder.weight']
- This IS expected if you are initializing BertModel from the checkpoint of a model trai
ned on another task or with another architecture (e.g. initializing a BertForSequenceCla
ssification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model
that you expect to be exactly identical (initializing a BertForSequenceClassification mo
del from a BertForSequenceClassification model).
```

Q1: Fill out the `get_cls_token_for_doc` function to return the [CLS] embedding for the input string. The output should be a single 768-dimensional numpy vector (see `4.embeddings/BERT.ipynb` for converting between a pytorch tensor and a numpy object).

In [68]:
```
def get_cls_token_for_doc(model, string):
    inputs = tokenizer(string, return_tensors="pt")
    # your code goes here

    #getting the tokens
    tokenizer.convert_ids_to_tokens(inputs["input_ids"][0])
```

```
    outputs = model(**inputs)

    #last_hidden_states = outputs.last_hidden_state

    return outputs.last_hidden_state[0].detach().numpy()[2]
```

In [69]:
```
bert_cls_data=get_document_reps_for_data(data, get_cls_token_for_doc, model)
```

Q2: Use these representations to find the episode that is most similar to the description "The one where they bounce back in time" by having the highest cosine similarity between representations. A sample function shell `run_query` is provided below, along with the only arguments you need, but feel free to adapt it as you see fit.

In [71]:
```
query="The one where they bounce back in time"
```

In [82]:
```
def run_query(query, data_with_reps, sequence_embedding_function, model):
    # your code goes here
    vals = []

    for eps in data_with_reps:
        comp_rep = sequence_embedding_function(model, query)
        cos_sim = cosine_similarity(eps[3], comp_rep)
        vals.append((cos_sim, query, eps[0]))

    for c, q, s in reversed(sorted(vals)):
        print("%.3f\t%s\t%s" % (c, q, s))
```

In [83]:
```
run_query(query, bert_cls_data, get_cls_token_for_doc, model)
```

```
0.396   The one where they bounce back in time  62
0.377   The one where they bounce back in time  111
0.373   The one where they bounce back in time  75
0.369   The one where they bounce back in time  92
0.366   The one where they bounce back in time  44
0.365   The one where they bounce back in time  81
0.350   The one where they bounce back in time  65
0.348   The one where they bounce back in time  32
0.347   The one where they bounce back in time  1
0.347   The one where they bounce back in time  118
0.345   The one where they bounce back in time  102
0.342   The one where they bounce back in time  150
0.333   The one where they bounce back in time  5
0.332   The one where they bounce back in time  96
0.330   The one where they bounce back in time  30
0.328   The one where they bounce back in time  36
0.323   The one where they bounce back in time  46
0.318   The one where they bounce back in time  91
0.318   The one where they bounce back in time  88
0.315   The one where they bounce back in time  121
0.312   The one where they bounce back in time  132
0.307   The one where they bounce back in time  55
0.304   The one where they bounce back in time  87
0.302   The one where they bounce back in time  149
0.302   The one where they bounce back in time  119
```

0.300 The one where they bounce back in time 74
0.299 The one where they bounce back in time 139
0.299 The one where they bounce back in time 4
0.298 The one where they bounce back in time 130
0.297 The one where they bounce back in time 11
0.296 The one where they bounce back in time 127
0.295 The one where they bounce back in time 59
0.294 The one where they bounce back in time 86
0.294 The one where they bounce back in time 66
0.294 The one where they bounce back in time 131
0.294 The one where they bounce back in time 58
0.290 The one where they bounce back in time 73
0.290 The one where they bounce back in time 8
0.287 The one where they bounce back in time 15
0.287 The one where they bounce back in time 112
0.287 The one where they bounce back in time 77
0.286 The one where they bounce back in time 138
0.286 The one where they bounce back in time 21
0.286 The one where they bounce back in time 9
0.285 The one where they bounce back in time 10
0.284 The one where they bounce back in time 24
0.284 The one where they bounce back in time 12
0.283 The one where they bounce back in time 128
0.283 The one where they bounce back in time 89
0.282 The one where they bounce back in time 100
0.280 The one where they bounce back in time 33
0.279 The one where they bounce back in time 70
0.278 The one where they bounce back in time 84
0.278 The one where they bounce back in time 25
0.277 The one where they bounce back in time 141
0.277 The one where they bounce back in time 122
0.276 The one where they bounce back in time 137
0.275 The one where they bounce back in time 72
0.273 The one where they bounce back in time 95
0.273 The one where they bounce back in time 90
0.270 The one where they bounce back in time 117
0.269 The one where they bounce back in time 41
0.269 The one where they bounce back in time 106
0.268 The one where they bounce back in time 79
0.268 The one where they bounce back in time 98
0.268 The one where they bounce back in time 152
0.267 The one where they bounce back in time 82
0.266 The one where they bounce back in time 69
0.266 The one where they bounce back in time 145
0.264 The one where they bounce back in time 146
0.264 The one where they bounce back in time 31
0.263 The one where they bounce back in time 71
0.263 The one where they bounce back in time 103
0.259 The one where they bounce back in time 68
0.259 The one where they bounce back in time 29
0.258 The one where they bounce back in time 148
0.257 The one where they bounce back in time 27
0.256 The one where they bounce back in time 123
0.256 The one where they bounce back in time 67
0.254 The one where they bounce back in time 126
0.254 The one where they bounce back in time 124
0.252 The one where they bounce back in time 26
0.252 The one where they bounce back in time 80
0.251 The one where they bounce back in time 54
0.251 The one where they bounce back in time 136
0.250 The one where they bounce back in time 147

0.249 The one where they bounce back in time 115
0.249 The one where they bounce back in time 125
0.248 The one where they bounce back in time 52
0.247 The one where they bounce back in time 28
0.247 The one where they bounce back in time 34
0.247 The one where they bounce back in time 108
0.246 The one where they bounce back in time 83
0.245 The one where they bounce back in time 35
0.243 The one where they bounce back in time 53
0.242 The one where they bounce back in time 47
0.240 The one where they bounce back in time 50
0.240 The one where they bounce back in time 116
0.240 The one where they bounce back in time 113
0.240 The one where they bounce back in time 42
0.239 The one where they bounce back in time 143
0.239 The one where they bounce back in time 99
0.239 The one where they bounce back in time 40
0.238 The one where they bounce back in time 57
0.235 The one where they bounce back in time 120
0.235 The one where they bounce back in time 78
0.233 The one where they bounce back in time 17
0.232 The one where they bounce back in time 37
0.231 The one where they bounce back in time 43
0.227 The one where they bounce back in time 51
0.224 The one where they bounce back in time 105
0.224 The one where they bounce back in time 85
0.223 The one where they bounce back in time 151
0.222 The one where they bounce back in time 104
0.221 The one where they bounce back in time 56
0.221 The one where they bounce back in time 19
0.219 The one where they bounce back in time 60
0.218 The one where they bounce back in time 76
0.217 The one where they bounce back in time 20
0.216 The one where they bounce back in time 140
0.216 The one where they bounce back in time 107
0.215 The one where they bounce back in time 16
0.212 The one where they bounce back in time 7
0.210 The one where they bounce back in time 3
0.210 The one where they bounce back in time 14
0.209 The one where they bounce back in time 110
0.208 The one where they bounce back in time 129
0.207 The one where they bounce back in time 94
0.206 The one where they bounce back in time 114
0.206 The one where they bounce back in time 63
0.205 The one where they bounce back in time 93
0.203 The one where they bounce back in time 101
0.196 The one where they bounce back in time 23
0.193 The one where they bounce back in time 45
0.192 The one where they bounce back in time 18
0.191 The one where they bounce back in time 142
0.190 The one where they bounce back in time 49
0.189 The one where they bounce back in time 48
0.187 The one where they bounce back in time 109
0.186 The one where they bounce back in time 39
0.182 The one where they bounce back in time 61
0.181 The one where they bounce back in time 64
0.179 The one where they bounce back in time 97
0.178 The one where they bounce back in time 133
0.171 The one where they bounce back in time 6
0.170 The one where they bounce back in time 134
0.167 The one where they bounce back in time 13

```
0.162   The one where they bounce back in time  2
0.143   The one where they bounce back in time  22
0.130   The one where they bounce back in time  135
0.124   The one where they bounce back in time  144
0.105   The one where they bounce back in time  38
```

Now let's try a sentence embedding model that was optimized for generating sentence representations: Sentence-BERT (Reimers and Gurevych 2019). Example usage (in the context of the Huggingface transformers library) can be found here.

In [84]:
```
sentence_model = SentenceTransformer('sentence-transformers/all-distilroberta-v1')
```

Q3: Fill out the `get_sentence_embedding` function below to return the sentence embedding for the input string, and use it again to find the episode that is most similar to the description "The one where they bounce back in time" by having the highest cosine similarity between representations. Which method for generating sentence embeddings appears better for this task?

In [105…
```
def get_sentence_embedding(model, string):
    # your code goes here
    embeddings = sentence_model.encode(string)
    return embeddings
```

In [106…
```
get_sentence_embedding(model, query)
```

Out[106…
```
array([ 2.21291883e-03,  1.46007640e-02,  2.03430578e-02,  1.27137015e-02,
        1.35774920e-02,  3.19657773e-02, -7.94165395e-03,  5.24583161e-02,
        2.90296637e-02, -1.41363693e-02,  4.94237766e-02, -2.31884580e-04,
       -4.03048992e-02, -6.15077540e-02,  9.52694193e-03, -7.15107704e-03,
       -6.67671785e-02, -4.34239171e-02, -1.10399239e-02, -2.21695676e-02,
       -1.34791841e-03, -3.53775956e-02, -7.93617591e-02,  8.38638619e-02,
       -3.64556462e-02,  1.69570819e-02, -8.85298103e-02,  1.07683558e-02,
       -8.18700530e-03,  3.06761004e-02,  1.76261924e-02, -1.28581850e-02,
        2.35316176e-02,  6.67219236e-02,  5.61144166e-02,  3.27455346e-03,
       -4.85332571e-02, -9.01158492e-04,  6.27200492e-03, -5.37139736e-03,
        1.40093248e-02,  1.00574549e-03,  2.96798851e-02,  1.23534417e-02,
       -1.32677779e-02, -6.48900867e-03,  1.86410528e-02,  4.17390689e-02,
       -1.89139359e-02, -3.09142452e-02,  3.27500366e-02,  2.01369170e-02,
        3.56953824e-03, -3.77495885e-02,  1.06609203e-02,  5.49859665e-02,
       -1.41724269e-03, -1.52485780e-02, -1.37812691e-02,  1.48049816e-02,
        2.49714293e-02, -9.09428224e-02, -2.07985006e-03,  2.63621733e-02,
        3.11674410e-03,  2.14692149e-02, -1.30433282e-02, -7.28110597e-02,
        1.96592454e-02, -1.46719180e-02,  1.44420527e-02,  2.45864019e-02,
       -4.54400778e-02,  2.89853720e-04,  7.60145532e-03,  3.64723196e-03,
       -1.69990733e-02, -6.93664700e-02, -1.38491075e-02,  2.56028473e-02,
```

-3.50530520e-02, 7.26699308e-02, -4.05667834e-02, 1.06763688e-03,
 4.28503491e-02, -6.79032430e-02, -1.14855748e-02, -1.70884803e-02,
 9.46743507e-03, 2.67948955e-02, -9.88203473e-03, 4.40861844e-02,
 5.08677177e-02, 3.89164500e-02, -3.79183963e-02, -9.36778076e-03,
-2.02575196e-02, 1.25979796e-01, -1.92892291e-02, 5.68098156e-03,
-4.03489619e-02, 5.32163270e-02, -5.44315018e-03, 5.55830076e-02,
-7.20151141e-03, 1.84596481e-03, -3.76855652e-03, 2.94402204e-02,
 2.28764433e-02, 3.08500882e-02, -4.25108112e-02, 2.08726246e-02,
-3.65397800e-03, 5.86143807e-02, -1.33511070e-02, -1.16881272e-02,
 2.20061373e-02, -2.04156861e-02, -3.02557778e-02, 3.20698917e-02,
 2.08282471e-02, -1.23807220e-02, 4.24493626e-02, 3.32105085e-02,
 7.04018101e-02, 9.06893332e-03, -4.06467430e-02, -1.51359066e-01,
-3.29957567e-02, -8.76825154e-02, 1.20234238e-02, -4.67043854e-02,
 4.51379009e-02, 7.87090417e-03, -1.95024759e-02, 3.00189126e-02,
 4.17987891e-02, 1.60717387e-02, 4.48036101e-03, 3.45836878e-02,
 3.99757698e-02, 2.72923447e-02, -1.91768585e-03, 1.79524515e-02,
 3.38673145e-02, -3.37349288e-02, 1.51583888e-02, 2.06432026e-02,
-2.84546819e-02, -5.17852344e-02, 8.12441483e-02, -1.81690678e-02,
 2.23133969e-03, 4.92140017e-02, -5.06111374e-03, 1.92196853e-02,
-4.71705347e-02, 9.11280606e-03, 3.81658785e-02, 4.57479917e-02,
 1.16539588e-02, -1.53999710e-02, -1.48039078e-02, 1.64523441e-02,
-3.35796946e-03, -4.39754650e-02, -3.30474451e-02, -8.44774942e-04,
 6.27329051e-02, -2.32955087e-02, -2.20720116e-02, 4.58791927e-02,
 5.94791099e-02, -5.52829504e-02, 3.45610492e-02, 6.18197396e-02,
-5.13229556e-02, 8.08027610e-02, 2.34802328e-02, 3.61699909e-02,
-4.68614744e-03, -2.40609725e-03, 7.38980100e-02, -5.79677196e-03,
-1.63228661e-02, 1.58600044e-02, -3.31827104e-02, -2.26014368e-02,
-4.26743664e-02, -2.03661229e-02, 4.15895320e-02, -3.00418194e-02,
-2.63435598e-02, -1.28771281e-02, -2.58608758e-02, -6.30370453e-02,
 9.14851651e-02, -1.95028745e-02, 1.56084625e-02, -1.98379196e-02,
 1.71796791e-02, 1.22316694e-02, 1.09745778e-01, -1.12986630e-02,
 4.90177125e-02, -1.59105342e-02, -1.33282430e-02, 5.45077436e-02,
-1.35276113e-02, -1.49595123e-02, 1.20830687e-03, -4.54155961e-03,
-1.98911913e-02, -2.28482541e-02, -2.25817561e-02, 3.55753466e-03,
-2.80248076e-02, 8.04722607e-02, -1.69929937e-02, -1.18243219e-02,
 7.42609054e-02, -3.13200988e-02, -2.71249246e-02, -3.02734300e-02,
-5.73949069e-02, 1.61248166e-02, -1.76638830e-02, 4.97527719e-02,
-1.86391233e-03, 3.01872119e-02, -5.42424023e-02, 5.21208532e-02,
 3.47083695e-02, 8.34543444e-03, 2.89557818e-02, -4.73210029e-03,
 1.00078307e-01, -2.02751048e-02, -6.53888285e-03, -1.58774480e-02,
-2.55581457e-02, -4.82113939e-03, 2.49504037e-02, -5.64011298e-02,
 1.60764754e-02, -1.76009759e-02, -1.66858104e-03, 3.18602063e-02,
 2.87729315e-02, 1.61234401e-02, 4.71981131e-02, 3.60711776e-02,
 8.81029107e-03, -4.05638739e-02, 3.06993891e-02, -2.26861779e-02,
-2.86178105e-02, 1.46398861e-02, 9.10236873e-03, 6.47793263e-02,
-9.16156732e-03, 3.49191763e-02, -3.98512296e-02, 2.17506830e-02,
-3.80389877e-02, 2.51834132e-02, 2.37333756e-02, 1.13705471e-02,
-4.66702394e-02, -2.15412546e-02, -4.15026024e-03, -1.34915877e-02,
 3.03957798e-02, 1.65170282e-02, -4.68902588e-02, 4.30448763e-02,
 4.75014932e-03, -2.10781526e-02, -9.00529325e-03, -7.51913711e-03,
-3.50393308e-03, -3.08090951e-02, -1.25842728e-02, 1.97249874e-02,
-3.02418899e-02, 2.49120388e-02, 1.95122194e-02, -1.48757761e-02,
-8.45433958e-03, 2.07474772e-02, 4.29975204e-02, 3.06963157e-02,
-2.70412583e-02, 2.15281337e-03, -1.35459034e-02, -5.44551685e-02,
-2.00507045e-02, 7.40589714e-03, 4.53561923e-04, -3.56789082e-02,
-4.48385105e-02, -4.19514365e-02, 5.66280410e-02, 3.45417224e-02,
-3.31271850e-02, -5.91354333e-02, -2.94692256e-02, 1.45531455e-02,
-1.29651614e-02, -2.08167769e-02, -4.14324440e-02, -7.38277275e-04,
-8.37282650e-03, 5.14540374e-02, -7.97048733e-02, 5.63222857e-04,
 1.38438931e-02, -8.54356389e-04, -1.02657499e-02, 7.20910877e-02,
 2.64527444e-02, 4.63393927e-02, -5.44629944e-03, -2.24378761e-02,

```
-6.05147444e-02, -8.12144857e-03,  3.84929497e-03,  3.33278328e-02,
 1.48666631e-02,  2.95839664e-02, -1.04039289e-01, -2.54411660e-02,
-3.08515970e-02,  2.64406819e-02, -2.93540582e-02,  6.08844534e-02,
 1.52755696e-02, -2.47605890e-02,  3.36184545e-04,  3.19134742e-02,
-4.39078249e-02,  1.57307629e-02,  1.04721906e-02, -6.71743508e-03,
-3.31027294e-03, -6.52832910e-02,  2.69277804e-02,  1.78861199e-03,
 4.97643016e-02,  4.67903614e-02, -3.02304495e-02, -2.36002225e-02,
-2.11849064e-03, -3.95075791e-03, -6.50295382e-03, -1.50645478e-02,
-2.40596390e-04,  1.52773233e-02,  3.91860958e-03, -2.34858394e-02,
 1.75554713e-03, -1.51093747e-03,  2.17265170e-02, -3.52287851e-02,
 1.72481798e-02, -5.95338494e-02, -4.09956351e-02, -6.63001230e-03,
 7.63502419e-02,  1.01583241e-03, -7.55239604e-03,  1.41998539e-02,
-1.46027403e-02,  3.80894728e-02,  2.47641969e-02, -2.05197316e-02,
 6.24312870e-02, -8.97597745e-02,  1.36640323e-02, -1.64970178e-02,
 5.14223166e-02, -3.51534002e-02, -1.45202391e-02,  2.31130440e-02,
-2.37241816e-02,  1.81850139e-02,  5.46971336e-03, -4.16800603e-02,
 9.78655517e-02, -1.81357097e-02,  2.49483678e-02, -2.88970657e-02,
-3.74293490e-03,  6.91632256e-02,  9.33621917e-03,  1.82613228e-02,
 3.82086150e-02,  2.88215522e-02,  6.31359965e-02, -5.21778911e-02,
 5.33160567e-02, -2.54744161e-02,  2.70235687e-02,  1.54799288e-02,
-4.76785526e-02,  5.17638624e-02, -1.59045160e-02, -1.64057184e-02,
 3.79515179e-02,  7.65309529e-03,  9.12372023e-03,  6.43153489e-02,
 1.27697969e-02,  5.90448491e-02,  1.05696190e-02,  4.63568280e-03,
-1.77777316e-02, -8.73347744e-03,  2.29637343e-02,  7.95334801e-02,
 4.97696549e-02,  2.25280114e-02,  3.95001546e-02,  3.02420463e-03,
 1.80288237e-02,  2.56689684e-03, -1.29529685e-02, -1.27854105e-03,
-7.42591918e-03, -2.24011764e-02, -2.48855017e-02, -8.55687484e-02,
-5.85518638e-03, -3.30785438e-02,  1.11911877e-03, -5.49122021e-02,
 4.05883193e-02,  2.52610967e-02,  1.92099754e-02, -4.84749489e-02,
-1.82878878e-02, -3.54616344e-02,  5.12124151e-02,  4.21108939e-02,
-1.46670556e-02, -1.53476857e-02,  6.14334792e-02,  3.49272713e-02,
 3.27575356e-02,  1.67912350e-03, -7.55056888e-02, -1.45413158e-02,
-1.32042579e-02, -5.17149568e-02, -2.84248814e-02, -2.16986239e-02,
-3.08782863e-03,  3.47218998e-02,  1.38619170e-03,  3.78282666e-02,
 2.83214240e-03, -1.75256226e-02, -1.62327848e-02,  9.29554552e-03,
 1.39869740e-02, -4.91586886e-03, -1.27024734e-02,  5.08511774e-02,
 2.84793340e-02,  2.01693531e-02, -2.07730308e-02, -1.24611938e-02,
-4.24916372e-02, -4.03589010e-02, -3.96966795e-03, -1.12628322e-02,
-9.84559581e-02, -3.65020894e-02,  3.17138359e-02,  1.71383545e-02,
-1.38486652e-02, -1.92760099e-02,  1.92616116e-02, -1.44866873e-02,
-1.22793848e-02,  3.09865773e-02, -2.00249758e-02,  4.56785709e-02,
 2.67754328e-02, -1.33322980e-02,  3.19032334e-02,  2.64700856e-02,
-1.30730476e-02, -3.52753839e-03,  2.17156510e-33,  1.33138550e-02,
-6.32691607e-02, -5.39007001e-02,  8.55311304e-02, -3.79473157e-02,
 6.71645254e-03, -4.49737310e-02,  7.68038929e-02, -1.82718318e-02,
-1.45877227e-02, -4.57998272e-03, -4.13810369e-03, -4.84352894e-02,
-1.85312815e-02,  2.96622496e-02,  4.62089293e-02, -5.45271393e-03,
-4.27619554e-02, -2.26776628e-03, -7.45747704e-03,  3.03296335e-02,
 7.30959419e-03, -3.76255140e-02,  2.93690450e-02,  1.10472729e-02,
 3.24803330e-02,  2.54544634e-02, -8.85757431e-03,  1.32278968e-02,
-3.85051593e-02, -7.65438098e-03,  8.48209951e-03,  4.03504185e-02,
 8.13582074e-03,  8.61408375e-03, -5.66714481e-02, -2.90653631e-02,
 1.69131458e-02, -5.38208173e-04,  2.04201452e-02, -5.79386542e-04,
-2.68766545e-02,  3.19870259e-03, -2.54053008e-02,  3.18425372e-02,
-3.37387510e-02, -3.89852971e-02, -1.87378339e-02, -3.57849523e-02,
 8.93797167e-03, -1.23559637e-02,  5.30179963e-02, -9.12249554e-03,
 6.24284893e-03, -4.70225215e-02,  3.70354876e-02, -6.30526096e-02,
-3.80523615e-02, -1.43564194e-02, -8.69926662e-05, -1.60269008e-03,
 4.75796163e-02,  3.00323078e-03,  5.53539507e-02, -3.72805223e-02,
-2.27514841e-02,  5.51960664e-03, -1.16593288e-02,  1.18770134e-02,
-5.34966253e-02, -3.12041733e-02, -2.53740009e-02, -1.18508544e-02,
```

```
 3.17209847e-02,  2.41814051e-02,  1.74945872e-02, -5.58640435e-02,
-4.69565811e-03, -2.11711954e-02,  1.17525766e-02, -3.19689922e-02,
-1.41299926e-02,  1.79970302e-02,  5.25570028e-02,  3.70890764e-03,
 3.91017720e-02,  3.32126804e-02, -3.53420153e-02, -6.92135915e-02,
 4.45570238e-02,  3.84343080e-02,  4.53064851e-02,  3.14874351e-02,
-4.38765965e-05,  1.99763384e-02, -9.83823277e-03, -5.66110872e-02,
-5.35560818e-03,  2.28485130e-02,  2.36424543e-02,  2.74402928e-02,
-3.58455218e-02, -5.33464327e-02,  1.65300667e-02,  4.65191714e-03,
-7.90528953e-03, -9.17639732e-02,  6.47867518e-03, -6.85673431e-02,
-3.51829119e-02, -1.01479264e-02,  7.02062175e-02,  3.16270487e-03,
 1.72238797e-02, -4.37075458e-02, -4.09999024e-03,  1.74359996e-02,
-2.13136040e-02, -8.76731984e-03,  3.55058610e-02,  5.91446571e-02,
-1.74465440e-02,  3.10921557e-02, -3.56930234e-02,  7.54021853e-02,
 2.39817072e-02, -1.43353995e-02, -4.20640446e-02,  3.74372415e-02,
 5.21604940e-02, -3.17335650e-02,  2.33216751e-02,  1.02511495e-01,
 1.40924603e-02,  8.06194991e-02,  2.54230406e-02,  3.54625061e-02,
-3.01836245e-02,  1.40863471e-02, -6.63911998e-02, -6.05957769e-02,
-4.75138240e-03,  1.26640638e-03,  7.68455677e-03, -1.69133488e-02,
-6.21553808e-02,  6.56201541e-02,  5.58134727e-02, -7.45311156e-02,
 1.21466238e-02,  1.41489552e-02, -6.20018654e-02,  9.46553797e-02,
 2.27337517e-02, -2.48843320e-02,  1.13459947e-02,  8.44585337e-03,
-9.08971578e-02, -4.53343056e-02,  2.83331480e-02,  6.48867488e-02,
 1.71370171e-02,  2.89198346e-02, -3.13544832e-02, -2.93983202e-02,
 2.54119895e-02, -4.29376401e-02, -3.70732130e-04, -4.01982293e-02,
-2.23859493e-02, -1.43441539e-02,  3.32331322e-02, -1.36592612e-01,
 2.10027490e-02,  3.45640369e-02, -2.24993401e-03,  2.95645036e-02,
-2.94183362e-02,  7.56334662e-02,  4.42030206e-02, -6.86599836e-02,
-3.15298028e-02, -5.26868254e-02, -3.67746279e-02,  2.23003961e-02,
-5.77544421e-02,  2.40107290e-02,  4.00874950e-02, -2.44222209e-03,
 3.21263596e-02, -5.11016436e-02, -4.71268734e-03,  4.81284633e-02,
-2.35034097e-02, -1.31728360e-02, -1.04371291e-02, -1.75945032e-02,
 1.85968529e-03,  2.49489211e-02, -3.04331002e-03,  2.60488000e-02,
 2.57156733e-02, -1.70342624e-02, -6.66107088e-02,  1.63565725e-02,
 3.75571032e-03,  2.32964684e-03,  1.29806595e-02, -2.93646497e-03,
 8.54911841e-03,  9.20998398e-03, -1.05059324e-02, -1.95238553e-02,
-3.26132141e-02,  5.25490846e-03, -2.80559703e-04,  1.46071287e-02,
-4.10063751e-03,  1.46662667e-02,  3.24111283e-02,  1.15096429e-03,
 2.27696523e-02, -5.82251325e-02,  2.30602436e-02,  9.00164992e-03,
-9.11707000e-04,  4.68654893e-02,  3.94302234e-03, -3.11784148e-02,
-3.03201471e-02, -6.22763997e-03, -1.36993630e-02, -7.15379640e-02,
-2.83833966e-02, -4.58457768e-02,  4.19994956e-03,  9.53249075e-03,
-4.10476513e-03, -2.92242263e-02,  3.37043102e-03, -5.71409427e-03,
-4.49105576e-02,  5.04756812e-03, -2.82579064e-02, -1.36668226e-02,
 1.31686712e-02, -2.64742263e-02,  1.83658360e-03, -7.73750171e-02,
 6.81264549e-02,  6.01057969e-02, -2.66325986e-03,  1.95708573e-02,
 6.41155392e-02, -1.23819280e-02,  1.39256716e-02,  1.46918604e-02,
-2.29183193e-02,  7.19881104e-03, -1.75621156e-02, -1.85609125e-02,
-4.86978143e-02, -2.84883601e-04, -3.20848599e-02,  8.21653008e-03,
-8.27786922e-02,  4.78361845e-02,  6.14723284e-03,  2.46668793e-02,
-1.25666236e-04, -4.68023643e-02,  4.28869911e-02,  2.85773873e-02],
dtype=float32)
```

In [94]:

```
def get_document_reps_for_data(data, sequence_embedding_function, model):

    # This function applies the sequence_embedding_function argument (a function itself
    # element in the input data list, and returns a copy of that list with an embedding

    data_with_reps=[]

    for episode, title, summary in data:
```

```
        data_with_reps.append((episode, title, summary, sequence_embedding_function(mod

    return data_with_reps
```

In [107…

```
sentence_transformer_data=get_document_reps_for_data_new(data, get_sentence_embedding,
```

In [97]:

```
def get_document_reps_for_data_new(data, sequence_embedding_function, model):

    # This function applies the sequence_embedding_function argument (a function itself
    # element in the input data list, and returns a copy of that list with an embedding

    data_with_reps=[]

    for episode, title, summary in data:
        data_with_reps.append((episode, title, summary, sequence_embedding_function(sen

    return data_with_reps
```

In [109…

```
query="The one where they bounce back in time"
run_query(query, sentence_transformer_data, get_sentence_embedding, sentence_model)
```

```
0.349   The one where they bounce back in time  91
0.341   The one where they bounce back in time  76
0.335   The one where they bounce back in time  143
0.334   The one where they bounce back in time  82
0.330   The one where they bounce back in time  47
0.325   The one where they bounce back in time  17
0.321   The one where they bounce back in time  83
0.313   The one where they bounce back in time  121
0.306   The one where they bounce back in time  123
0.304   The one where they bounce back in time  29
0.303   The one where they bounce back in time  84
0.293   The one where they bounce back in time  15
0.292   The one where they bounce back in time  130
0.290   The one where they bounce back in time  43
0.289   The one where they bounce back in time  98
0.285   The one where they bounce back in time  27
0.283   The one where they bounce back in time  13
0.281   The one where they bounce back in time  92
0.278   The one where they bounce back in time  59
0.277   The one where they bounce back in time  6
0.277   The one where they bounce back in time  99
0.277   The one where they bounce back in time  18
0.275   The one where they bounce back in time  104
0.274   The one where they bounce back in time  54
0.274   The one where they bounce back in time  74
0.273   The one where they bounce back in time  52
0.272   The one where they bounce back in time  120
0.270   The one where they bounce back in time  8
0.268   The one where they bounce back in time  22
0.267   The one where they bounce back in time  131
0.266   The one where they bounce back in time  61
0.265   The one where they bounce back in time  23
0.265   The one where they bounce back in time  68
0.264   The one where they bounce back in time  62
0.261   The one where they bounce back in time  89
```

0.261 The one where they bounce back in time 117
0.260 The one where they bounce back in time 81
0.260 The one where they bounce back in time 10
0.260 The one where they bounce back in time 133
0.258 The one where they bounce back in time 5
0.257 The one where they bounce back in time 57
0.257 The one where they bounce back in time 151
0.255 The one where they bounce back in time 67
0.254 The one where they bounce back in time 9
0.254 The one where they bounce back in time 140
0.248 The one where they bounce back in time 66
0.247 The one where they bounce back in time 55
0.247 The one where they bounce back in time 48
0.247 The one where they bounce back in time 73
0.244 The one where they bounce back in time 19
0.243 The one where they bounce back in time 58
0.243 The one where they bounce back in time 87
0.242 The one where they bounce back in time 124
0.242 The one where they bounce back in time 146
0.240 The one where they bounce back in time 103
0.240 The one where they bounce back in time 16
0.240 The one where they bounce back in time 32
0.240 The one where they bounce back in time 41
0.239 The one where they bounce back in time 37
0.237 The one where they bounce back in time 64
0.237 The one where they bounce back in time 80
0.237 The one where they bounce back in time 12
0.237 The one where they bounce back in time 79
0.236 The one where they bounce back in time 126
0.234 The one where they bounce back in time 105
0.234 The one where they bounce back in time 71
0.233 The one where they bounce back in time 21
0.232 The one where they bounce back in time 2
0.231 The one where they bounce back in time 38
0.229 The one where they bounce back in time 39
0.229 The one where they bounce back in time 56
0.224 The one where they bounce back in time 102
0.221 The one where they bounce back in time 112
0.220 The one where they bounce back in time 88
0.220 The one where they bounce back in time 90
0.219 The one where they bounce back in time 49
0.218 The one where they bounce back in time 97
0.217 The one where they bounce back in time 119
0.217 The one where they bounce back in time 69
0.217 The one where they bounce back in time 25
0.215 The one where they bounce back in time 144
0.214 The one where they bounce back in time 1
0.213 The one where they bounce back in time 110
0.210 The one where they bounce back in time 106
0.207 The one where they bounce back in time 31
0.207 The one where they bounce back in time 11
0.206 The one where they bounce back in time 65
0.206 The one where they bounce back in time 96
0.205 The one where they bounce back in time 116
0.204 The one where they bounce back in time 20
0.204 The one where they bounce back in time 50
0.204 The one where they bounce back in time 46
0.203 The one where they bounce back in time 107
0.202 The one where they bounce back in time 33
0.201 The one where they bounce back in time 109
0.197 The one where they bounce back in time 34

```
0.197   The one where they bounce back in time  86
0.197   The one where they bounce back in time  36
0.194   The one where they bounce back in time  100
0.192   The one where they bounce back in time  142
0.191   The one where they bounce back in time  134
0.188   The one where they bounce back in time  108
0.188   The one where they bounce back in time  115
0.187   The one where they bounce back in time  35
0.187   The one where they bounce back in time  114
0.186   The one where they bounce back in time  70
0.185   The one where they bounce back in time  135
0.185   The one where they bounce back in time  45
0.184   The one where they bounce back in time  141
0.181   The one where they bounce back in time  3
0.180   The one where they bounce back in time  127
0.179   The one where they bounce back in time  113
0.179   The one where they bounce back in time  101
0.177   The one where they bounce back in time  53
0.174   The one where they bounce back in time  148
0.172   The one where they bounce back in time  118
0.171   The one where they bounce back in time  28
0.171   The one where they bounce back in time  94
0.170   The one where they bounce back in time  44
0.170   The one where they bounce back in time  24
0.169   The one where they bounce back in time  26
0.168   The one where they bounce back in time  30
0.168   The one where they bounce back in time  145
0.167   The one where they bounce back in time  147
0.167   The one where they bounce back in time  95
0.157   The one where they bounce back in time  125
0.153   The one where they bounce back in time  7
0.152   The one where they bounce back in time  14
0.150   The one where they bounce back in time  129
0.149   The one where they bounce back in time  132
0.148   The one where they bounce back in time  78
0.143   The one where they bounce back in time  77
0.141   The one where they bounce back in time  4
0.140   The one where they bounce back in time  60
0.138   The one where they bounce back in time  111
0.137   The one where they bounce back in time  152
0.136   The one where they bounce back in time  85
0.131   The one where they bounce back in time  42
0.131   The one where they bounce back in time  139
0.128   The one where they bounce back in time  122
0.127   The one where they bounce back in time  93
0.120   The one where they bounce back in time  72
0.115   The one where they bounce back in time  138
0.112   The one where they bounce back in time  137
0.107   The one where they bounce back in time  128
0.103   The one where they bounce back in time  149
0.095   The one where they bounce back in time  150
0.089   The one where they bounce back in time  51
0.079   The one where they bounce back in time  63
0.077   The one where they bounce back in time  40
0.050   The one where they bounce back in time  136
0.035   The one where they bounce back in time  75
```

In [ ]: