

CIS 5930 : Term Project

Sajib, Fatema, Arunima



Text summarization is the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks).

Text Summarization

— Page 1, [Advances in Automatic Text Summarization](#), 1999.

ARTICLE: McNair was a regular in the first team squad under Louis van Gaal last season, while Love made two appearances under the Dutchman. Neither of the defensive duo played under Moyes, but he was aware of their presence in the youth team and hopes to persuade them to continue their development in the North East. A joint fee of around £8 million is thought to have been agreed with United and it is now up to the players to decide if they want to leave Old Trafford. Sunderland, though, are confident both [...]

SUMMARY: *David* Moyes **has returned** to **former club Manchester** United to **strengthen his** Sunderland squad **after agreeing** a fee **for Paddy** McNair and **Donald** Love.

Open news sites, and
start reading every news?

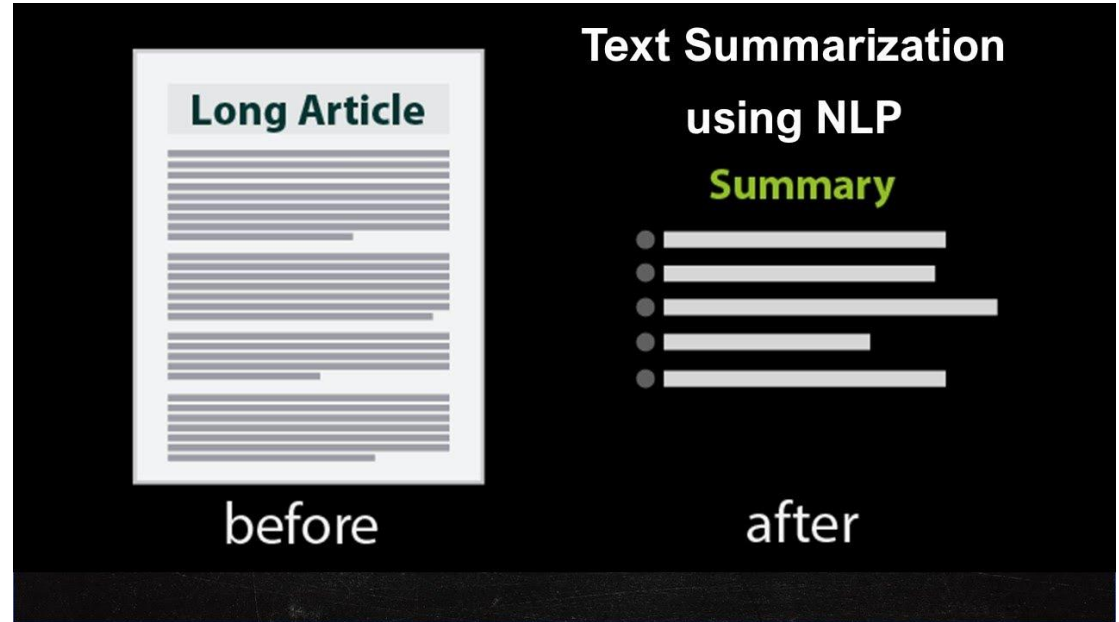
Probably not.

Glance at the headlines
and decide!



Summaries can be created automatically using ML!

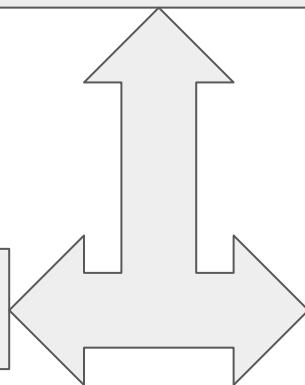
The google news and other news aggregator apps are doing it already!

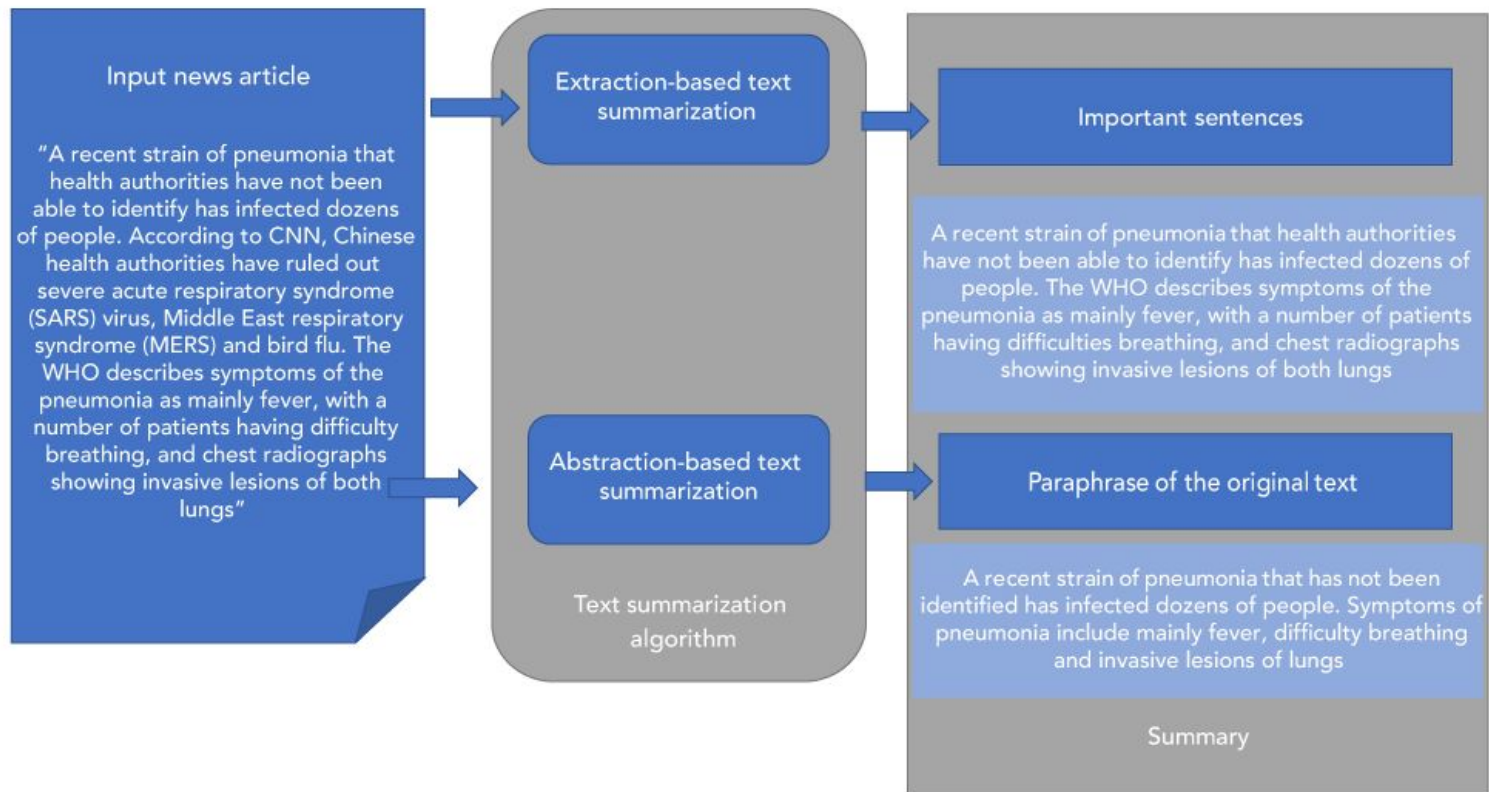


Text Summarization

Extractive Summarization

Abstractive Summarization







NEWSROOM

SUMMARIZATION DATASET

CORNELL NEWSROOM is a large dataset for training and evaluating summarization systems. It contains 1.3 million articles and summaries written by authors and editors in the newsrooms of 38 major publications. The summaries are obtained from search and social metadata between 1998 and 2017 and use a variety of summarization strategies combining *extraction* and *abstraction*.

bbc.com

Coverage: 0.9

Density: 3

Compression: 10:1

SUMMARY: Bradford City and Port Vale play out a goalless draw on the opening day of the League One season.

ARTICLE: Neither Bradford City nor day of the League One season in a Valley Parade. But Bradford's Stu performance than opposite number match without being able to supply came to scoring was in the 12th mi

Read Full Article

theguardian.com

Coverage: 0.62

Density: 1.12

Compression: 52:1

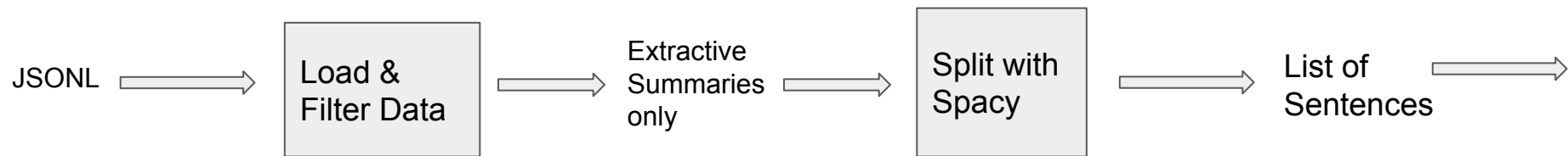
SUMMARY: A series of **sceptical** exhibitions in Lisbon **questions** the **purpose** of architecture, **writes** **Rowan Moore**

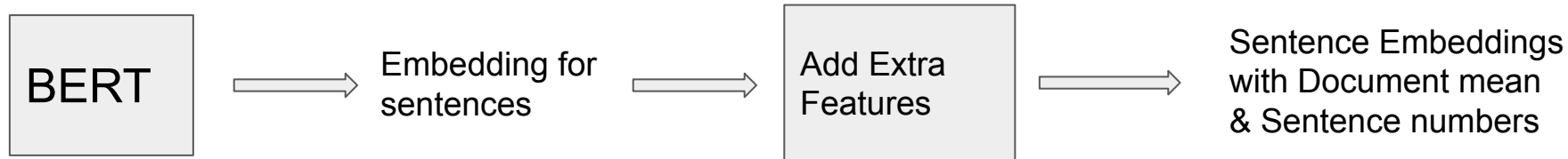
ARTICLE: Redefining space: The Nation Room - Embassy of No Land by Paulo Moreira and Kiluanji Kia Henda, part of The Real and Other Fictions exhibition. I'm standing in the shadow of an equestrian statue in a Lisbon square, as part of a small crowd, listening to eager young architects explain their projects through megaphones. We're not meant to be here, as a special stage has been constructed nearby for this and other events. But the sun is hot, the stage has no shade and the pompous statue casts a [...]

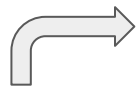
Read Full Article



Data Preprocessing Pipeline

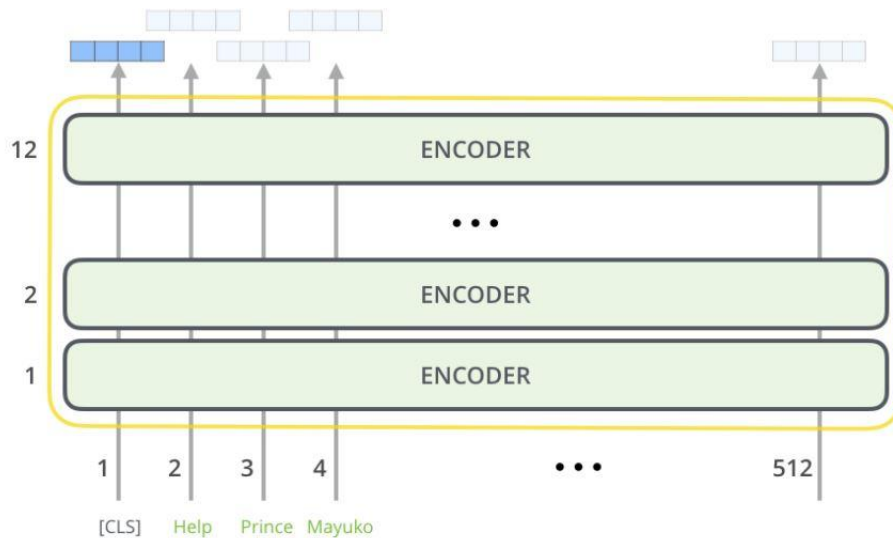






Embedding

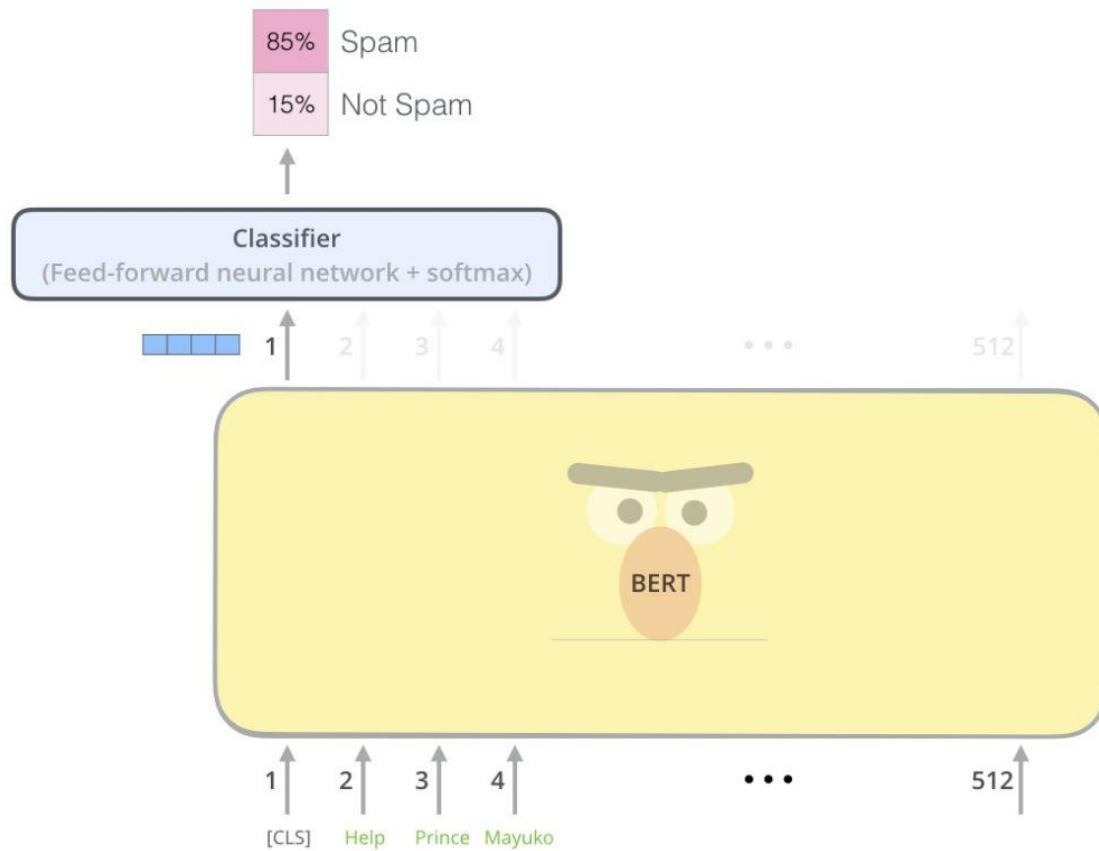
-0.34	-0.84	0.20	-0.26	-0.12	0.23	1.04	-0.16	0.31	0.06	0.30	0.33	-1.17	-0.30	0.03
-------	-------	------	-------	-------	------	------	-------	------	------	------	------	-------	-------	------

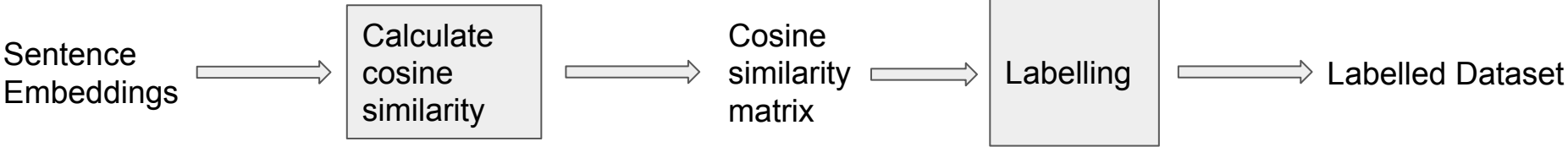


BERT

Sentence







988235
h
ractive
A.J. BENZA & MICHAEL LEWITTES\n\n, If Simon Rex l
0.708276, 0.4615584, 0.25829703, -1.157553, -0.673
Simon Rex looks a little familiar, it may not hav
0.050813835, -0.2324406, 0.03911153, -0.8688194, -
0.15, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
1, 3

5,000 Articles

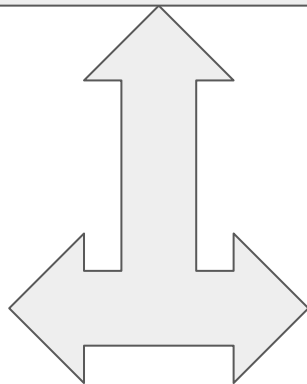
150,000 Sentences

Machine Learning Pipeline

Supervised Learning

Only Embedding
Information

Including Sequential
information

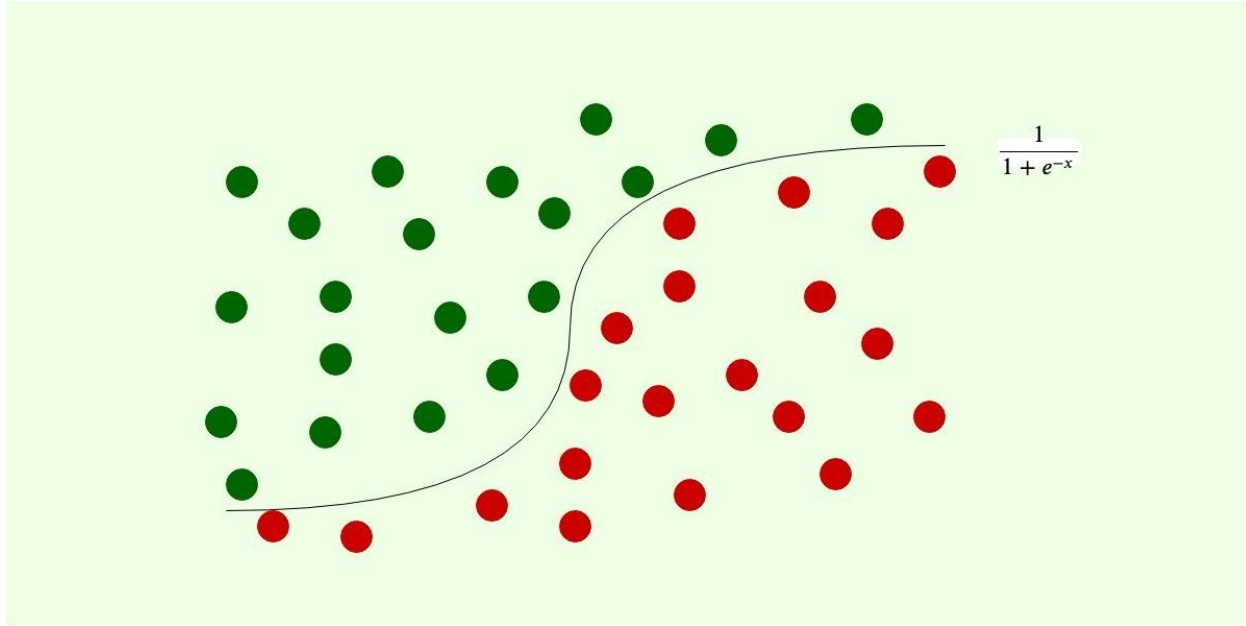


Supervised Learning Using Only Embedding Information

Logistic Regression

Neural Networks

Logistic Regression



Neural Network

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 50)	76850
dense_1 (Dense)	(None, 50)	2550
dense_2 (Dense)	(None, 1)	51

Total params: 79,451

Trainable params: 79,451

Non-trainable params: 0

Supervised Learning Using Sequential Information

LEDE3

Logistic Regression

LSTM

LEDE3

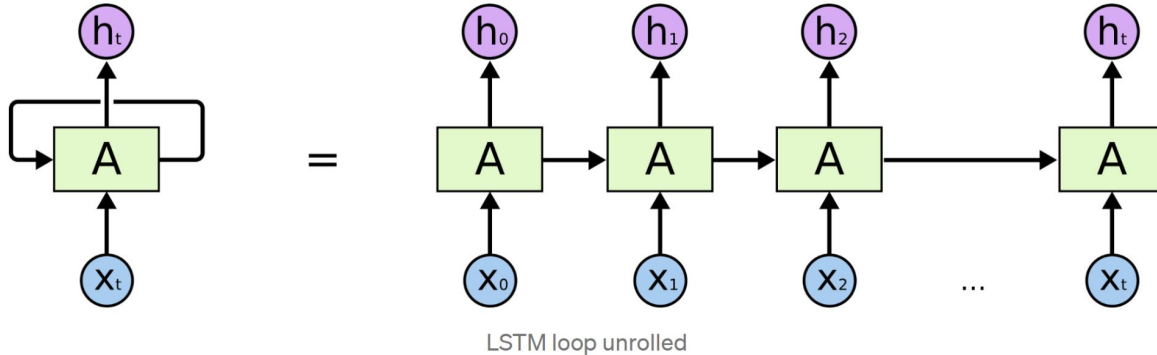
First three sentences of any article.

Exceedingly difficult baseline to beat for extractive summarization evaluated with ROUGE scores.

LSTM

Unlike standard feedforward neural networks, LSTM has feedback connections.

Keeps contextual information.



Metrics

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Recall

$$\text{Recall} = \frac{TP}{TP+FN}$$

From all the positive classes, how many we predicted correctly?

Precision

$$\textit{Precision} = \frac{TP}{TP+FP}$$

From all the classes we have predicted as positive, how many are actually positive?

F1_Score

$$F1Score = \frac{2 * precision * recall}{precision + recall}$$

Gives us Recall and Precision at the same time!!

Original: *"the quick brown fox jumps over"*

Unigrams: [*'the', 'quick', 'brown', 'fox', 'jumps', 'over'*]

Bigrams: [*'the quick', 'quick brown', 'brown fox', 'fox jumps', 'jumps over'*]

Trigrams: [*'the quick brown', 'quick brown fox', 'brown fox jumps', 'fox jumps over'*]

ROUGE-N measures the number of matching ‘n-grams’ between generated text and a ‘reference’.

Model Score

Supervised Learning With only Embedding Information

Neural Network (Dense Dense)

	ROUGE-1	f1	recall	precision
NN 25 25 Bal	0.403	0.379	0.577	
NN 25 50 Bal	0.396	0.390	0.551	
NN 50 50 Bal	0.415	0.397	0.583	

	ROUGE-1	f1	recall	precision
Logistic Reg	0.416	0.398	0.578	
NN 50 50 Bal	0.415	0.397	0.583	

Supervised Learning With Sequential Information

ROUGE-1 f1 recall precision

LSTM

LSTM Uni 25	0.596	0.579	0.778
LSTM Uni 50	0.596	0.576	0.783
LSTM Bi 25	0.596	0.575	0.781
LSTM Bi 50	0.599	0.577	0.785
LSTM Bi 75	0.599	0.577	0.785

ROUGE-1 f1 recall precision

LEDE3	0.589	0.588	0.757
LogReg Def	0.525	0.518	0.696
LSTM Bi 50	0.599	0.577	0.785

Example

Original

By CORKY SIEMASZKO DAILY NEWS STAFF WRITER Mother Teresa believed she was possessed by the Devil, the archbishop of Calcutta said yesterday. So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said. Archbishop Henry D'Souza's bizarre revelation came as millions yesterday marked the fourth anniversary of Mother Teresa's death. But D'Souza told CNN and The Associated Press in India he truly believed that.

Logistic Reg

Archbishop Henry D'Souza's bizarre revelation came as millions yesterday marked the fourth anniversary of Mother Teresa's death. The Catholic cleric said he diagnosed the demon in Mother Teresa shortly before she had a fatal heart attack Sept. 5, 1997, and died at age 87. Mother Teresa won a Nobel Prize for her life's work, and Pope John Paul has begun the process of declaring her a saint.

Neural Net

So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said. The Catholic cleric said he diagnosed the demon in Mother Teresa shortly before she had a fatal heart attack Sept. 5, 1997, and died at age 87. So did Michael Cuneo, author of "American Exorcism," which hits bookstores next week.

LSTM

BY CORKY SIEMASZKO DAILY NEWS STAFF WRITER With Emily GestThursday, September 6th 2001, 2:23AM Mother Teresa believed she was possessed by the Devil, the archbishop of Calcutta said yesterday. So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said.

LEDE3

BY CORKY SIEMASZKO DAILY NEWS STAFF WRITER With Emily GestThursday, September 6th 2001, 2:23AM Mother Teresa believed she was possessed by the Devil, the archbishop of Calcutta said yesterday. So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said.

Original

By CORKY SIEMASZKO DAILY NEWS STAFF WRITER Mother Teresa believed she was possessed by the Devil, the archbishop of Calcutta said yesterday. So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said. Archbishop Henry D'Souza's bizarre revelation came as millions yesterday marked the fourth anniversary of Mother Teresa's death. But D'Souza told CNN and The Associated Press in India he truly believed that.

Logistic Reg

Archbishop Henry D'Souza's bizarre revelation came as millions yesterday marked the fourth anniversary of Mother Teresa's death. The Catholic cleric said he diagnosed the demon in Mother Teresa shortly before she had a fatal heart attack Sept. 5, 1997, and died at age 87. Mother Teresa won a Nobel Prize for her life's work, and Pope John Paul has begun the process of declaring her a saint.

Neural Net

So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said. The Catholic cleric said he diagnosed the demon in Mother Teresa shortly before she had a fatal heart attack Sept. 5, 1997, and died at age 87. So did Michael Cuneo, author of "American Exorcism," which hits bookstores next week.

LSTM

BY CORKY SIEMASZKO DAILY NEWS STAFF WRITER With Emily GestThursday, September 6th 2001, 2:23AM Mother Teresa believed she was possessed by the Devil, the archbishop of Calcutta said yesterday. So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said.

LEDE3

BY CORKY SIEMASZKO DAILY NEWS STAFF WRITER With Emily GestThursday, September 6th 2001, 2:23AM Mother Teresa believed she was possessed by the Devil, the archbishop of Calcutta said yesterday. So the revered nun, whom the Vatican hopes to make a saint, underwent an exorcism and afterward "slept like a baby," he said.

THANK

YOU !

