

Analyzing Orphaned Triples in a Large-Scale Graph Database

CIS4914 – Final Report

Spring 2020

Gary Gurlaskie (g.gurlask@ufl.edu)

Individual Project

16 April 2020

Abstract

This project emphasized data analysis and visualization, and was focused on UF VIVO, an 7.1 GB RDF graph database that documents UF research efforts. The goal was to identify data in VIVO that had become "orphaned" (disconnected) from the graph.

The analysis was performed ad hoc using Python. Analysis performed included examining zero- or low-degree nodes, orphaned connected components, and performing constraint validation against the published ontology.

The work yielded over 10,000 orphaned triples. However, the quality of the data was high overall; 94% of the graph was connected in a robust way.

Introduction

a := 1;

```
lin114-00:3% java network 1337
Received: Packet0, 1, PASS
Received: ACK0, DROP
Received: Packet0, 1, DROP
Received: Packet0, 1, DROP
Received: Packet0, 1, PASS
Received: ACK0, DROP
Received: Packet0, 1, PASS
Received: ACK0, PASS
Received: Packet1, 2, PASS
Received: ACK1, DROP
Received: Packet1, 2, PASS
Received: ACK1, PASS
Received: Packet0, 3, DROP
Received: Packet0, 3, PASS
Received: ACK0, PASS
Received: Packet1, 4, DROP
Received: Packet1, 4, PASS
Received: ACK1, CORRUPT
Received: Packet1, 4, DROP
Received: Packet1, 4, PASS
Received: ACK1, CORRUPT
Received: Packet1, 4, PASS
Received: ACK1, PASS
Received: Packet0, 5, CORRUPT
Received: ACK1, DROP
Received: Packet0, 5, PASS
Received: ACK0, DROP
Received: Packet0, 5, CORRUPT
Received: ACK0, PASS
Received: Packet1, 6, CORRUPT
Received: ACK0, DROP
Received: Packet1, 6, CORRUPT
Received: ACK0, PASS

lin114-01:~/test> java receiver lin114-00 1337
Waiting 1, 1, 1 0 289 The, ACK0
Waiting 1, 2, 1 0 289 The, ACK0
Waiting 1, 3, 1 0 289 The, ACK0
Waiting 0, 4, 2 1 541 quick, ACK1
Waiting 0, 5, 2 1 541 quick, ACK1
Waiting 1, 6, 3 0 552 brown, ACK0
Waiting 0, 7, 4 1 333 fox, ACK1
Waiting 0, 8, 4 1 333 fox, ACK1
Waiting 0, 9, 4 1 333 fox, ACK1
Waiting 0, 10, 5 0 646 jumped, ACK1
Waiting 1, 11, 5 0 646 jumped, ACK0
Waiting 1, 12, 5 0 646 jumped, ACK0
Waiting 1, 13, 6 1 445 over, ACK0
Waiting 1, 14, 6 1 445 over, ACK0
Waiting 1, 15, 6 1 445 over, ACK0
Waiting 0, 16, 6 1 444 over, ACK1
Waiting 0, 17, 6 1 444 over, ACK1
Waiting 1, 18, 7 0 321 the, ACK0
Waiting 1, 19, 8 1 449 lazy, ACK0
Waiting 1, 20, 8 1 449 lazy, ACK0
Waiting 0, 21, 8 1 448 lazy, ACK1
Waiting 0, 22, 8 1 448 lazy, ACK1
Waiting 0, 23, 9 0 361 dog., ACK1
Waiting 0, 24, 9 0 361 dog., ACK1
Waiting 1, 25, 9 0 360 dog., ACK0
Message: The quick brown fox jumped over the lazy dog.
lin114-01:~/test>

lin114-02:3% java sender lin114-00 1337 message.txt
Waiting ACK0, 1, DROP, resend Packet0
Waiting ACK0, 2, DROP, resend Packet0
Waiting ACK0, 3, DROP, resend Packet0
Waiting ACK0, 4, DROP, resend Packet0
Waiting ACK0, 5, ACK0, send Packet1
Waiting ACK1, 6, DROP, resend Packet1
Waiting ACK1, 7, ACK1, send Packet0
Waiting ACK0, 8, DROP, resend Packet0
Waiting ACK0, 9, ACK0, send Packet1
Waiting ACK1, 10, DROP, resend Packet1
Waiting ACK1, 12, DROP, resend Packet1
Waiting ACK1, 14, ACK1, send Packet0
Waiting ACK0, 15, DROP, resend Packet0
Waiting ACK0, 16, DROP, resend Packet0
Waiting ACK0, 17, ACK0, send Packet1
Waiting ACK1, 18, DROP, resend Packet1
Waiting ACK1, 19, ACK0, resend Packet1
Waiting ACK1, 20, DROP, resend Packet1
Waiting ACK1, 21, DROP, resend Packet1
Waiting ACK1, 22, DROP, resend Packet1
Waiting ACK1, 23, DROP, resend Packet1
Waiting ACK1, 25, ACK1, send Packet0
Waiting ACK0, 26, ACK0, send Packet1
Waiting ACK1, 27, DROP, resend Packet1
Waiting ACK1, 28, ACK0, resend Packet1
Waiting ACK1, 29, DROP, resend Packet1
Waiting ACK1, 30, ACK1, send Packet0
Waiting ACK0, 31, DROP, resend Packet0
Waiting ACK0, 32, ACK1, resend Packet0
Waiting ACK0, 33, DROP, resend Packet0
Waiting ACK0, 34, ACK0, no more packets to send
lin114-02:4%
```

Figure 1: An example execution.

Acknowledgments

The author thanks Christopher Barnes for advising the project, and the UF CTS-IT for providing access to the VIVO data. The author also thanks Taeber Rapczak for his excellent help and guidance.

Biography

Gary Gurlaskie is an undergraduate student studying Mathematics and Computer Science at the University of Florida. He worked at Ultimate Software (data platform engineering) and Intel (software engineering). He plans to graduate in May 2020, and upon graduation will move to Portland to work at Intel full-time. He has always wanted to have a cat, so he will obtain one when he moves to Oregon.