COLX 523 Week 1

Work distribution & Checklist

[Toshiko] Set up a private repo for the group

[Toshiko] Teamwork contract

[Andrew] Project proposal

[Toshiko] Compelling proposal

[Yuesheng, Jiang] Corpus Collection POC (proof-of-concept)

[Team] Brainstorming

Timelines

W1 & 2: Web scraping at least 1 million words
W3: Annotations (annotating a fraction of the corpus is okay)
W4: Web interface

Brainstorming

The source of data: Twitter

Annotations: We first annotate named entities (names of crypto-currency, exchange, etc.), then annotate semantic relationships between those entities (pick 2 or 3 predicates from DBpedia)

Topic: Crypto-currency: NER (Cryptocurrency/Coinbase/bitcoin/NFT)