CS/INFO 6742 – Natural language processing and social interaction
Lecture 3 handout, Sep 2, 2021

About the A1 datasets

Slashdot	ChangeMyView (Reddit)
Not explicitly task-oriented; generic news	Focused on persuasion tasks
Peer-moderation tags --- noisy "ground truth" for comment aspects	Original-poster deltas: "ground truth" for persuasion success
Peer-moderator scores range from -2 to +5	Upvotes/downvotes not supposed to be used. In arbitrary Reddit, an obfuscated difference is displayed.
	Pairings of successful and unsuccessful threads for the same original post

Slashdot data:
1. Comment IDs and parent-of-comment IDs (<parent>)
2. <modclass>, <modscore>
3. 3 versions:
    a. htmltext: quoting structure *sometimes* made explicit, e.g., <i> or <div class="quote">
    b. tokenized (<tokenext> (?)). Example: "n't".
    c. sentence-tokenized. (with newlines)

Features – mentioned/displayed from a ChangeMyView paper, but good for both datasets and lots of potential projects!