

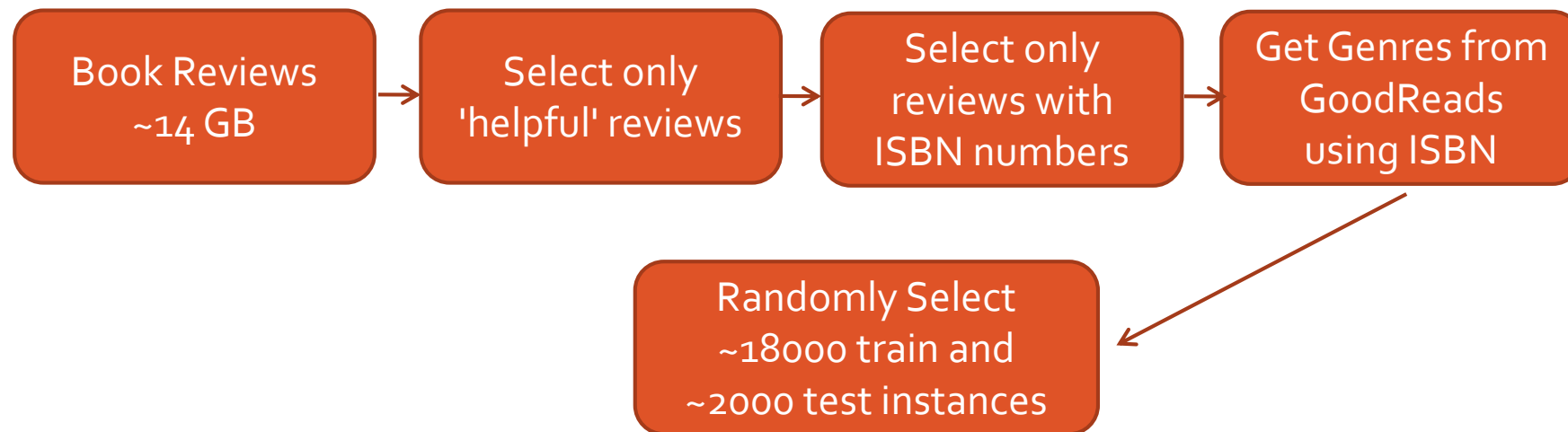
BOOK REVIEWS & GENRES

Antariksh Bothale and Maria Antoniak

LING 575 -- Spring 2014

Corpus Collection

- Amazon Book Review Corpus



Aspect Extraction

- We use MALLET's LDA model to extract topics for each sentence in each review.
- We would love to use seed words as in the Mukerjhee paper, but we could not find a package, and weren't sure if coding it from scratch in a short time line would be wise
- Maybe there is another method to get more specific results?

Aspect Extraction

- story line told plot reader telling moving turns compelling slow tale interesting lines twists moves mystery bottom pace quickly
- series books left readers entire leave volume rest disappointed leaves trilogy find wanting happened set waiting direction pick fill
- characters character main story plot development developed interesting lead cast descriptions realistic drawn strong dialogue personality believable setting intriguing
- book excellent guide good reference advice practical complete introduction study resource purpose skills fast comprehensive essential title tool serve
- part parts chapters major authors variety close wide longer lead range discuss broken individual subjects contrast themes similar discusses
- man woman young states girl finds united beautiful named heart lady tells sees friend protect meets runs determined mysterious
- read book easy understand follow fun quick difficult enjoyable put easier helped helps pick entertaining full skip format fairly
- point view points starting position perspective views argument critical challenge fair support generally alternative arguments sides balanced ultimately offer
- writing style funny written prose humor entertaining engaging writer narrative author insightful wit brilliant voice makes witty tone clever

Genre Merging

- We scraped book reviews from GoodReads
- User-defined and classified, lots of over-specific genres (Mermaids, Satanism, Sex Work and so on)
- Can't expect to discover them via plain LDA, and so manually merging them into 20-ish broad genres

What next?

- Finish genre merging (too many genres)
- Use the aspects to cluster the book reviews into genres.
- Maybe reverse our task and use genres to better extract aspects.