

Text Analytics

LING 223 / DIGI 223

How can we understand the vast amount of human knowledge being stored as text data?

Text Analytics is about deriving meaning from written documents. In this course you will learn how to analyze data from millions of documents – far too many to read them each individually. From novels to news articles, speeches to social media posts, movie subtitles to product reviews, you will learn how to collect diverse datasets and use them to answer questions from a large range of perspectives.

Example Topics

Gender.

We know it is possible to identify a writer's gender. Is it possible to identify the genders of characters within novels? Do female authors represent male characters more "accurately" than male authors represent female characters? What types of novels have the most "accurate" male and female dialogue?

Politics.

We know it is possible to track the salience of political issues from political speeches and Twitter posts. Do these two data sources agree about which issues are important? Is their agreement consistent across nations? What issues do non-politicians most talk about that are ignored by politicians?

Geography.

We know it is possible to characterize places like Christchurch from tourism reviews, news articles, and Wikipedia. How consistent are these characterizations across populations? For example, do Japanese and Australian tourists have the same perceptions of Christchurch? Which types of tourists are the most similar? Which other tourist destinations are the most similar to Christchurch?

Culture.

We know that cultures and languages frame events in different ways. Is it possible to understand the differences between cultures given competing news accounts of a single set of events? Which cultures share a similar framing of events? Can competing frames be used to explain choices about which details or which participants to highlight?