Trinity Term (May 2016) Supervisor: Dr. CARL VOGEL

TRINITY COLLEGE DUBLIN

Abstract

Computer Science Department
School Of Computer Science & Statistics

Integrated Masters In Computer Science

Narrative Entity & Relation Alignment: Structuralist Analysis In Natural Language Processing

by Neil Hyland

"Structuralist Analysis" is a linguistic approach to understanding text, pioneered by the French anthropologist Claude Levi-Strauss in the 1960s. It aims to represent the structure of a text (e.g. a narrative or story) in a formal manner: a table of data points describing major events in the order they occurred (the rows) and their shared themes (the columns).

The research undertaken for this work was to devise and implement a possible set of algorithms or procedures to perform structuralist analysis computationally, with the desired output being matrices describing the structure of narrative Elements (events occurring in the text) and Entities (people/places/things etc…). This was achieved with a four-stage process: where Elements were generated from the text (1), then processed using the K-Means statistical clustering algorithm to approximate the aforementioned thematic grouping (2), Entities were generated from mentions in the text (3), and then all Element and Entity data is combined into an output form of a two-dimensional matrix in the style of Lévi-Strauss' original table approach, and a three-dimensional spatial matrix representing the position and alignment of Entities relative to the Elements they were found in (4).

The findings of this work can be summarised as follows: computationally implementing structuralist analysis is feasible, and with further development could yield a new and different approach to modelling text data for the purposes of comparison, aggregation, summarisation or other use.