Natural Language Generation with Markov Chains and Grammar

Sam Zhang
Computer Systems Lab 2009-2010

Abstract

How do we assign meaning to words? This project investigates semantics from a lexical perspective, using the WordNet and OpenCyc ontologies to create a semiotic map of our consensus reality. Given a list of words, how can we find the word least like the others? Through a heuristicial search across the hypernym ontology, computational semantics can discover the contextual meaning of words, even when the only context given is the other words from which it must differentiate itself. This method, which has not been given a name previously, will hitherto be known as dynamic word sense disambiguation.
Language can be generated stochastically using Markov Chain databases. This project explores the use thereof in junction with the aforementioned semantic web.

Background and Introduction

To create a comprehensible text, this project uses a corpus-based method to gleam the basic grammatical rules and vocabulary, then using those, to piece together a new text (Figures 1 and 2).

Discussion

As can be seen, the two inaugural addresses (Figures 1 and 2) reflect the semiotic landscape of the time. Bush uses “power”, “burns”, “faiths”, “democratic”, “free”, and “idealistic”. Washington uses “nor party animosities”, reflecting his anti-partisan views; “peculiarly conscious”, “surely pledges”, and other phrases uncommon in contemporary political discourse.

Results and Conclusions

The project successfully regenerates texts using a second-order Markov Chain, but it may benefit from a future adding of a context-free grammar to impose further structure on the work. Thusly the text may not end abruptly with misplaced punctuation, although that problem was much more common with a first-order Markov Chain than now.

Figure 1: 2005 Bush's inaugural address:

In the truths of Sinai , the questions that come to those who feel its power , it burns those who feel its power , it burns those who feel its power , and the varied faiths of our soldiers . You have seen together . For as long as whole regions of the United States to seek and support the growth of democratic movements and institutions in every generation by reaffirming all that is the concentrated work of helping raise up free governments K the idealistic work of generations . The great objective of ending tyranny is the...

Figure 2: 1789 Washington's inaugural address:

Than side no local prejudices or attachments , no separate views nor party animosities , will misdirect the comprehensive and equal eye which ought to be peculiarly conscious of his own deficiencies . In these honorable qualifications I behold the surest pledges that as on one side no local prejudices or attachments , no separate views nor party animosities , will misdirect the comprehensive and equal eye which ought to be suppressed ...