

2011/2012 Seminar Series

[www.cs.unb.ca/
seminarseries](http://www.cs.unb.ca/seminarseries)



From DNA to Jay-Z: How Ideas from Bioinformatics can Automate Finding Rhymes in Rap Music

By:

Dr. Dan Brown
Cheriton School of Computer Science
University of Waterloo

Unlike most kinds of music, the core of rap music is found in the rhythm and rhyme of its lyrics. Different artists or subgenres will use different kinds of rhyme, which in some cases can be extremely complicated: the end of one line may rhyme with several parts of the previous line, and in some cases, rhymes may be imperfect.

Detecting these complex rhyme patterns manually is time consuming and tedious. We have designed a system for automatic rhyme annotation. Our approach is founded on several bioinformatics ideas. First, using a test corpus of known rhymes, we develop a probabilistic model of rhymed and unrhymed syllables. Then, we use that model to build a log-likelihood ratio scoring matrix for identifying what is and is not a rhyme. Finally, we create a local alignment procedure to find high-scoring lyrics segments.

Our procedure has high sensitivity and specificity in identifying true rhymes in an annotated corpus; essentially, it identifies most complex rhymes, and identifies few false rhymes. We can use it to characterize artists, and then to develop classifiers for individual artists with surprising success.

Joint work with MMath student Hussein Hirjee.

Dan Brown is Associate Professor of Computer Science at the University of Waterloo, where he has been since 2001. From 2000 to 2001, he worked on the human and mouse genome projects at the Whitehead/MIT Center for Genome Research. His interests are in algorithms for understanding the information in discrete sequences, particularly identifying patterns in DNA and protein sequences.

Friday, September 30th @ 10:30am
Gillin Hall, room C-111