

CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, April 21

2:00 – 3:00 PM

Maria Sanford, Room 101

***TEXT MINING: SOME COMPLEXITIES AND
TECHNIQUES OF ANALYZING NATURAL
LANGUAGE***

ROGER BILISOLY

CENTRAL CONNECTICUT STATE UNIVERSITY

ABSTRACT

Natural languages such as English are complex. Basic tasks such as identifying words pose more difficulties than one would suppose. But even given flawless word identification, word frequencies are typically very low, and sample sizes are never large enough to assume the limiting case is ever approximately true. Hence the usual statistical techniques are often inapplicable to analyzing text. This talk will first discuss some of these problems posed by natural languages, and then will outline some techniques to transform a text to a database, which then can be analyzed by either statistical or data mining techniques. Some of these techniques will be illustrated on literary texts, including Charles Dickens' *A Christmas Carol* and several short stories by Edgar Allan Poe.

AFTERMATH:

**Refreshments will follow the colloquium at Castaneda's
(1590 Stanley St. – across from the administration building)**

For further information:

gotchevi@ccsu.edu 860-832-2839
castanedan@ccsu.edu 860-832-2851