University of Toronto, Department of Computer Science
CSC 2501/485F—Computational Linguistics, Fall 2015

Reading assignment 2

Due date: On CDF at 13h10, Wednesday 7 October 2015.
Late write-ups will not be accepted without a valid excuse.
This assignment is worth 5% of your final grade.

Read this paper

Porter, Martin F. An algorithm for suffix stripping. *Program*, 14(3), 1980, 130–137. [Reprinted in: Sparck Jones, Karen and Willet, Peter (editors), *Readings in Information Retrieval*, San Francisco: Morgan Kaufmann, 1997, 313–316.]

What to write

Write a short summary of the paper, and a set of questions and points for discussion following your randomly-assigned position, following the general guidelines for write-ups.

Background and context: This paper is a *contrast* to what we will see in class regarding the morphological analysis of words in preparation for parser. It has a different purpose and, obviously, a very different approach. It assumes a task of matching documents to queries by simple string matching, and tries to improve that. The Cranfield 200 collection that it refers to is a standard test-set of documents and queries.

Errata

(1) In the table on the left-hand side of page 314, the headings *precision* and *recall* should be interchanged.

(2) There is a formatting error on page 316 in Step 5b. It should read as follows, with one rule and two examples:

(m > 1 and *d and *L) $\rightarrow$ single letter	controll	$\rightarrow$ control
	roll	$\rightarrow$ roll