

# Story Extraction by Information Distance Models in Large Tour Guide Document

Junnan CHEN (20595284), j486chen@uwaterloo.ca

April 3, 2015

## Introduction

The Internet makes available a tremendous amount of text that has been generated for human consumption. In most cases, only a small part of the large text is useful and needed. Unfortunately, this information is not easily manipulated or analyzed by computers. *Story extraction* (SE) is the process of filling fields in a databases by automatically extract story fragments of human-readable text. Example include extracting specific story of a person, a place and an object from a large document which contains redundant information.