**Language-Independent Multi-Document Text Summarization with Document-Specific Word Associations**

#### Abstract

propose a word association based method for generating summaries in a variety of languages.

#### Introduction

Useful, needed.

Language-independent：

the language only need to follow the two conditions:

the text can be split to sentences (based on punctuation)

and sentences further to words (based on white space)

Contributions:

Using document-specific word associations as a model of the document:

For instance, accident is a frequent word in news stories, and so is Obama at the time of writing of this paper. A hypothetical document talking about an accident to President Obama is characterized by the combination of these two common words, and our goal is to be able to recognize such unexpected combinations. In contrast, a purely keyword-based method fails to discover the connection, and may actually miss both words if they are sufficiently common in news in general.

Find the set of sentences as the summary of documents: greedy one and genetic algorithm

#### Related Work

Extraction-based summarization

Abstraction-based summarization

Hybrids

Unsupervised or supervised

The most successful approach before: UWB

#### Problem

Formulation: U, D, d

Evaluation: sim()

Complexity: can be reduced the set cover problem

#### Method

Document-specific word associations

Mixture model for word co-occurrence: express what is expected

Log-likelihood ratio: (not understand)

Select sentences

pick sentences that cover as many of the associations as possible

pick sentences that cover the most central nodes in the term-association graph

combine the two strategies above