Heading for the document

Your paragraph goes here, **Croatian Text Summarizer (CROSUM)  
  
Tomislava Lauc, Nives Mikelié, Damir Boras  
Docent, Research Assistant, Associate Professor  
University of Zagreb, Faculty of Philosophy, Department of Information Sciences, I. Luéiéa 3  
tlauc@fize.hr, nmikelic@ffe.hr, dboras@ffze hr  
  
It is a custom to speak of an extract of n%  
condensation of input text. More precisely, n%  
  
Abstract. The paper describes automatic  
summarization of the scientific papers in  
  
Croatian language. The goal of the CROSUM is  
to generate extracts with high percent of extract-  
worthiness and about the same as the  
author's abstract. This preliminary research  
shows that extracts generated using the  
lemmatized wordforms dictionary are not quite  
different from extracts that are given on the base  
of the non-lemmatized wordforms dictionary.  
  
The research brought us to conclusion that we  
should develop a technique for identifying cue  
phrases from training corpus or some linguistic  
  
   
  
   
  
technique in order to improve the text  
summarization for Croatian language.  
Keywords. Text summarizer, Croatian  
  
language, extract, inflected language, tfidf words  
1. Introduction  
  
Text summarization is the process of,  
distilling the most important information from a  
source in order to produce an abridged version of  
a text for a particular user and a task. Given a  
document, text summarization is concerned with  
the generation of a shorter version which  
preserves the meaning of the original text.  
  
The word summary is used in a variety of  
contexts: depending on the input, one can have  
single or multiple-document summaries, while  
depending on the output one can have extract or  
abstract-like summaries.  
  
So, one of the goals of the automatic text  
summarization is to automatically generate  
extracts by selecting salient sentences from the  
original text, An extract is therefore a summary  
consisting entirely of the material copied from  
the input, but the sentences selected are those  
sentences of a text that are the most  
representative of pertinent information. The  
success of such summarization system relies on  
the use of appropriate features to select the  
salient sentences.  
  
   
  
of the input’s words may appear in the extract, or  
1n% of the input’s sentences may appear in the  
extract, or even n% of the input paragraphs may  
appear in the extract.  
  
Unlike an extract, an abstract is a summary  
which re-phrases content coherently and also  
contains at least some of the material that is not  
present in the input, In general, abstracts offer  
the possibility of higher degrees of condensation:  
a short abstract may offer more information than  
a longer extract.  
  
Furthermore, depending on the usage, a  
summary can be indicative or informative. An  
indicative summary can provide only an  
indication of the main topics in the input text.  
Thus, an indicative abstract aims at helping the  
user to decide whether to read the information  
source, or not. By contrast, an informative  
abstract covers all the salient information in the  
source at some level of detail, ice, it can reflect  
to a certain extent the semantic content of the  
input text as well.  
  
Also, depending on the purpose, a summary  
can be generic. i, it can reflect the author's  
point of view with respect to all important topics  
in the input text, or it can be query oriented (also,  
user-focused or topic-focused), in other words, it  
can reflect only the topics in the input text that  
are specific to a given query.  
  
   
  
   
  
2. The system for CROatian text  
  
SUMmarization - CROSUM.  
  
Summarization system can have a lot of  
different parameters, such as: compression rate  
(summary length vs. source length), audience  
(user focused vs. generic), relation to source  
(extract vs. abstract), function (indicative  
informative), coherence (coherent  
incoherent), span (single- vs. multi- document  
summarization), language (monolingual,  
multilingual or cross-lingual), genre (special  
  
   
  
   
  
27th Int. Conf. Information Technology Interfaces ITI 2005, June 20-23, 2005, Cavtat, Croatia**, and *these words are italic*

## Heading level 2