Abstract:
With the continuing growth of the quantity of on-line text information, triggered in part by the growth of the World Wide Web, it is especially useful to have tools which can help users digest information content. Text summarization addresses this need by taking a source text, selecting the most important portions of it, and presenting coherent summary to the user in a manner sensitive to the user's or application's needs. The goal of this talk is to show how these objectives can be achieved through an efficient use of lexical cohesion. The current work addresses both generic and query-based summaries. I will present an approach for identifying the most important portions of the text which are topically best suited to represent the source text according to the author's views or in response to the user's interests. This identification must also take into consideration the degree of connectiveness among the chosen text portions so as to minimize the danger of producing summaries which contain poorly linked sentences. I will present a system that handles these objectives, discuss its performance, and compare it to other systems in the context of Document Understanding Conference (DUC) evaluation. Finally, I will outline some future works.