Chapter 10

New Formats and Interfaces for Multi-Document News Summarization and its Evaluation

Bettina Berendt
KU Leuven, Belgium

Mark Last
Ben-Gurion University of the Negev, Israel

Ilija Subašić
KU Leuven, Belgium

Mathias Verbeke
KU Leuven, Belgium

ABSTRACT

News production, delivery, and consumption are increasing in ubiquity and speed, spreading over more software and hardware platforms, in particular mobile devices. This has led to an increasing interest in automated methods for multi-document summarization. The authors start this chapter with discussing several new alternatives for automated news summarization, with a particular focus on temporal text mining, graph-based methods, and graphical interfaces. Then they present automated and user-centric frameworks for cross-evaluating summarization methods that output different summary formats and describe the challenges associated with each evaluation framework. Based on the results of the user studies, the authors argue that it is crucial for effective summarization to integrate the user into sense-making through usable, entertaining, and ultimately useful interactive summarization-plus-document-search interfaces. In particular, graph-based methods and interfaces may be a better preparation for people to concentrate on what is essential in a collection of texts, and thus may be a key to enhancing the summary evaluation process by replacing the “one gold standard fits all” approach with carefully designed user studies built upon a variety of summary representation formats.

DOI: 10.4018/978-1-4666-5019-0.ch010
1. INTRODUCTION

Automatic text summarization is a computer-aided process of distilling the most important information from one or several textual sources on a common topic. Based on the number of input sources, summarization methods are categorised into single-document and multi-document approaches. The output of most summarization methods is a natural-language text in itself.

One distinguishes between summarization by extraction and by abstraction (Hahn & Mani, 2000). Extractive methods select elements from the original text(s) and recompile them to form a summary. Abstractive methods build an internal semantic representation and then use natural-language generation techniques to create a summary. Due to the remaining limitations of natural-language generation methods, extractive summarization remains the dominant approach today.

In this chapter, we focus on extractive summarization of text streams, particularly news streams. In this domain, multi-document summarization is especially relevant because of the multitude of sources. Popular applications of these ideas are news aggregators such as Google News or Yahoo News. However, these do not aim at producing a coherently-seeming natural-language text, but a search-engine-type collation of information. This echoes the findings of Barzilay, McKeown, and Elhadad (1999) that pure extraction may be inappropriate in multi-document summarization especially of news texts, because it may produce summaries which are overly verbose or biased towards some sources. In contrast to the static methods of multi-document summarization, STORIES (Subašić & Berendt, 2010a) is designed for both summarization and search in a time-indexed collection of documents. STORIES uses story graphs to represent summaries.

The classical summary evaluation methods suffer from severe limitations, especially when applied to Temporal Text Mining techniques, which may produce summaries in both textual and graph formats. Hence, we present an automated and a user-centric framework for the cross-evaluation of news summaries. An initial user study of the proposed frameworks provides encouraging results. We use these results to outline directions for future work.

The contribution of this chapter is threefold: First, we give an overview of key concepts and formats of text summarization in general and news summarization in particular (Section 2). Second, we present a critical re-assessment of evaluation methodology. In Section 3, we discuss the issue of evaluation of summarization methods, highlighting in particular the difficulties of cross-evaluating text-based as well as the emerging multitude of non-text-based summaries. We also take a closer look at a form of human input that is crucial for all automated evaluations: the “ground truth” summaries and their construction.

We argue that current practice severely limits our ability to evaluate automatic methods and that this calls for interactive, graphical approaches. Third, in Sections 4 and 5 we propose a new approach to evaluating summarization methods. We describe the automated and the user-centric frameworks for cross-evaluating methods that output different summary formats, and we describe the challenges associated with each evaluation framework. In Section 6, we draw conclusions from these findings and outline key elements of future multi-document summarization methods.

2. KEY CONCEPTS AND FORMATS IN TEXT SUMMARIZATION

Extractive summarization aims at the selection of a subset of the most relevant fragments from a source text into the summary. The fragments can be paragraphs (Salton, Singhal, Mitra, & Buckley, 1997), sentences (Luhn, 1958), keyphrases (Turney, 2000; Litvak, Aizenman, Gobits, Last, & Kandel, 2011) or keywords (Litvak & Last, 2008).