

# Chapter XVI

## Automatic Text Summarization in Digital Libraries

**Shiyan Ou**

*University of Wolverhampton, UK*

**Christopher S. G. Khoo**

*Nanyang Technological University, Singapore*

**Dion Hoe-Lian Goh**

*Nanyang Technological University, Singapore*

### ABSTRACT

*This chapter describes various text summarization techniques and evaluation techniques that have been proposed in literature and discusses the application of text summarization in digital libraries. First, it introduces the history of automatic text summarization and various types of summaries. Next, it reviews various approaches which have been used for single-document and multidocument summarization. Then, it describes the major evaluation approaches for assessing the generated summaries. Finally, it outlines the principal trends of the area of automatic text summarization. This chapter aims to help the reader to obtain a clear overview of the text summarization field and facilitate the application of text summarization in digital libraries.*

### INTRODUCTION

With the rapid growth of the World Wide Web, more and more information is available and accessible online and information overload becomes a big problem for users. Automatic text summa-

rization has attracted attention both in research communities and commercial organizations as a solution for reducing information overload and helping users to scan a large number of documents to identify documents of interest. Text summarization is regarded as “the process of distilling

the most important information from a source to produce an abridged version for a particular user and task” (Mani & Maybury, 1999, p. ix). It is an important function that should be available in large digital library systems, information retrieval systems, and Web search engines, where the retrieval of too many documents and the resulting information overload is a major problem for users.

In addition to facilitating the user’s searching and browsing, text summarization can do much more in digital libraries. Rather than serving as information providers, digital libraries are becoming knowledge repositories which strive to add value to the collections that they create and maintain. Text summarization is expected to be most helpful in this aspect since it could perform knowledge integration and support knowledge discovery and knowledge acquisition, especially in a multidocument environment.

## **BACKGROUND AND TYPES OF SUMMARIES**

Research in automatic text summarization has had a history of almost 50 years since the earliest attempt by Luhn (1958). However, there was little work and slow progress in the first 30 years. In the 1990s, as a result of information explosion in the World Wide Web, automatic text summarization became crucial to reduce information overload and this brought about its renaissance. It could be used for different purposes and different users and, thus, various types of summaries have been constructed.

Depending on the summarization method, a summary can be an *extract* (produced by sentence extraction) or an *abstract* (produced by an abstraction process). Using Mani’s (2001a, p. 6) definition:

- An extract is a summary consisting entirely of material copied from the input; and

- An abstract is a summary at least some of whose material is not present in the input.

Since extracts are much easier to be constructed automatically than abstracts which require more complex techniques such as rephrasing and paraphrasing, extracts are generally used in current digital library systems.

With reference to the content and intended use, a summary can be *indicative*, *informative*, or *evaluative* (Borko & Bernier, 1975):

- An indicative summary provides an indication of what the original document is about. It can help users to determine whether the original document is worth reading or not, but users have to consult the original for details.
- An informative summary reflects the content of the original document and represents the content in a concise way. It can be used as a substitute for the original document so that users do not need to read the original.
- An evaluative or critical summary not only contains the main topics of the original document but also provides the abstractor’s comments on the document content.

Indicative summaries are more generally used in current digital library systems to help users identify documents of interest. On the other hand, informative summaries are more often used for news articles to inform users about news events, for example, Columbia’s Newsblaster<sup>1</sup>.

Depending on the purpose and intended users, a summary can be *generic* or *user-focused* (Mani, 2001):

- A generic summary covers all major themes or aspects of the original document to serve a broad readership community rather than a particular group.
- A user-focused (or topic-focused, query-oriented) summary favors specific themes

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/automatic-text-summarization-digital-libraries/19879](http://www.igi-global.com/chapter/automatic-text-summarization-digital-libraries/19879)

## Related Content

---

### Automation Process in Nigerian Academic Libraries

Janet Onomeh Ubogu (2012). *International Journal of Digital Library Systems* (pp. 40-46).

[www.irma-international.org/article/automation-process-in-nigerian-academic-libraries/83501](http://www.irma-international.org/article/automation-process-in-nigerian-academic-libraries/83501)

### Does 3D Cellular Automata Using WordNet Can Improve Text Clustering?

Abdelmalek Amine, Reda Mohamed Hamouand Michel Simonet (2014). *International Journal of Digital Library Systems* (pp. 16-28).

[www.irma-international.org/article/does-3d-cellular-automata-using-wordnet-can-improve-text-clustering/141372](http://www.irma-international.org/article/does-3d-cellular-automata-using-wordnet-can-improve-text-clustering/141372)

### The Digital Library: A Multi-Faceted Information and Communication System

Wolfgang Ratzek (2012). *International Journal of Digital Library Systems* (pp. 62-70).

[www.irma-international.org/article/the-digital-library/99595](http://www.irma-international.org/article/the-digital-library/99595)

### Cross-Cultural Design and Usability of a Digital Library Supporting Access to Maori Cultural Heritage Resources

Chern L. Liew (2005). *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific* (pp. 285-297).

[www.irma-international.org/chapter/cross-cultural-design-usability-digital/8144](http://www.irma-international.org/chapter/cross-cultural-design-usability-digital/8144)

### Digital Libraries and the Role of Digital Librarians

Adetola Kehinde (2018). *Handbook of Research on Managing Intellectual Property in Digital Libraries* (pp. 98-119).

[www.irma-international.org/chapter/digital-libraries-and-the-role-of-digital-librarians/188545](http://www.irma-international.org/chapter/digital-libraries-and-the-role-of-digital-librarians/188545)