Abstract

Information is essential for all purpose of activities such as research, business decision making, etc. In this internet technology age there is no scarcity of information also. But if the information reveals the identity of a person or if it discloses confidential matters, then such information is a serious threat to privacy. So before publishing or sharing documents, the sensitive information should be removed or masked. This is the major goal of Text sanitization. Several semi-automatic and automatic methods are used for identifying sensitive information and thereby sanitizing the document by removing such terms. This broadens the users using the document due to their lowered classification level and also privacy is preserved.

References

A Review on Text Sanitization


Index Terms

Computer Science Information Sciences
Keywords
Document Declassification  Data Publishing  Term correlation  Privacy  Information
Theory
Named Entity Recognition