

## Chapter 10

# Implementation and Testing Details of Document Classification

### ABSTRACT

*It is trivial to achieve a recall of 100% by returning all documents in response to any query. Therefore, recall alone is not enough, but one needs to measure the number of non-relevant, for example by computing the precision. The analysis was performed for 30 documents to ensure the stability of precision and recall values. It is observed that the precision of large documents is less than a moderate length document, in the sense that some unimportant keywords get extracted. The reason for this may be attributed to the frequent occurrence and its unimportant role in the sentence.*

### SYSTEM TESTING

#### Reuters Data Set

Researchers have used benchmark data, such as the Reuters- 21578 corpus of newswire test collection (Sholom M. W., Indurkha, N., Zhang, T. and Damerau, F. 2010), to measure advances in automated text classification. We performed testing of our system using a sample of the same.

DOI: 10.4018/978-1-7998-3772-5.ch010

## **Modules of Execution**

1. Document Entry
2. Stop Word removal
3. Stemming
4. Keyword generation
5. Document Classification

## **Document Entry**

Doc\_id : DOC1

Doc\_content :

*Table 1. Words after tokenization*

hard
problem
text
classification
aspects
potential
solution
keyword
extraction
maximal
frequent
item
set
used
attributes
mining
association
rules
basis
measuring
similarity
new
documents
existing
association rules
issue
keyword
extraction
text
collection
emerging
research
filed
promotes
maximal
frequent
item
set
generation

9 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/implementation-and-testing-details-of-document-classification/268469](http://www.igi-global.com/chapter/implementation-and-testing-details-of-document-classification/268469)

## Related Content

---

### A Data Management Framework for Nurses Using E-Health as a Service (eHaaS)

Heng Liu, Rui Liu, Zhimei Liu, Xuena Han, Kaixuan Wang, Li Yang and FuGuo Yang (2023). *International Journal of Data Warehousing and Mining* (pp. 1-16).

[www.irma-international.org/article/a-data-management-framework-for-nurses-using-e-health-as-a-service-ehaas/319736](http://www.irma-international.org/article/a-data-management-framework-for-nurses-using-e-health-as-a-service-ehaas/319736)

### Shaping Content Strategies with User Analytics and Identities: How User Analytics is Shaping Editorial Strategy, Driving Marketing, and Generating New Revenue

Tricia Syed (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1209-1228).

[www.irma-international.org/chapter/shaping-content-strategies-with-user-analytics-and-identities/150212](http://www.irma-international.org/chapter/shaping-content-strategies-with-user-analytics-and-identities/150212)

### Digitization Initiatives and Knowledge Management: Institutionalization of E-Governance in Teaching, Learning and Research in East African Universities

A. M. Chailla, F. W. Dulle and A. W. Malekani (2009). *Social and Political Implications of Data Mining: Knowledge Management in E-Government* (pp. 288-301).

[www.irma-international.org/chapter/digitization-initiatives-knowledge-management/29077](http://www.irma-international.org/chapter/digitization-initiatives-knowledge-management/29077)

### Ooredoo Rayek: A Business Decision Support System Based on Multi-Language Sentiment Analysis of Algerian Operator Telephones

Badia Klouche, Sidi Mohamed Benslimane and Sakina Rim Bennabi (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1262-1279).

[www.irma-international.org/chapter/ooredoo-rayek/308543](http://www.irma-international.org/chapter/ooredoo-rayek/308543)

## Finding Persistent Strong Rules: Using Classification to Improve Association Mining

Anthony Scime, Karthik Rajasethupathy, Kulathur S. Rajasethupathy and Gregg R. Murray (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 28-49).

[www.irma-international.org/chapter/finding-persistent-strong-rules/73432](http://www.irma-international.org/chapter/finding-persistent-strong-rules/73432)