

ABSTRACT

SYSTEMS AND METHODS FOR DETECTING SIMILARITY OF DOCUMENTS

5

Disclosed are a system and a method for detecting a similarity of a first document with at least one second documents, comprising: extracting a text from the first document and the at least one second documents, respectively; segmenting the text of the first document into a first plurality of units and the text of the at least one
10 second document into a second plurality of units; and comparing the first plurality of segmented units with the second plurality of segmented units to determine whether the first document is similar to at least one of the at least one second document.

Figure 1

