

METHODS AND TEXT SUMMARIZATION SYSTEMS FOR DATA LOSS PREVENTION AND AUTOLABELLING

ABSTRACT

Methods and systems for data loss prevention and autolabelling of business categories and confidentiality based on text summarization are provided. The method for data loss prevention includes entering a combination of keywords and/or keyphrases and offline unsupervised mapping of a path of transfer of specific groups of documents. The offline unsupervised mapping includes keyword/keyphrase extraction from the specific groups of documents and normalization of candidates. The method further includes vectorization of the extracted keywords/keyphrases from the specific groups of documents and quantitative performance measurement of the keyword/keyphrase extraction to derive keywords and/or keyphrases suitable for data loss prevention.

FIG. 1