Analysis of Utilization of Biological Resources Using Text Mining Based on Freshwater Biodiversity Information Platform

Ji-Hae Lee[‡], Ye-seul Kwan[‡], Jungwook Park[‡], Sang Myeon Park[‡], Jeong Su Oh[‡]

‡ Nakdonggang National Institute of Biological Resources, Sangju-si, Gyeongsangbuk-do, South Korea

Corresponding author: Jeong Su Oh (ofang@nnibr.re.kr)

Abstract

The Nagoya Protocol on access to genetic resources and fair and equitable sharing of benefits arising from their utilization in the convention on biological diversity entered into force on October 12, 2014. Accordingly, attention toward securing the sovereignty and discovering the utilization value of biological resources has been increasing to secure national competitiveness. We are developing a freshwater biodiversity information platform for the systematic conservation and industrialization of freshwater biodiversity in South Korea. The platform comprises an integrated management system of freshwater bioresources for systematic registration and management of freshwater biodiversity information based on databases; a management system of storage for managing freshwater biological specimen; a utilization information system that manages efficacy, experimental method, and activity produced by the Nakdonggang National Institute of Biological Resources and external big data such as literature and patent; and a freshwater bioresources culture collection for preservation, ordering and deposition of biological resources. These systems are connected organically. Text mining, one of the big data technologies, helps to determine the utility of biological resources through comprehensive analysis. We tried to establish utilization foundations by predicting the usability of biological resources through systematic collection, processing, and analyzing external data, such as abstract, in order to support industrialization of national freshwater bioresources. Through

text mining, we constructed a literature-based corpus and preprocessed the corpus with lowercase conversion and removal of stop word. Then, a word cloud was created and statistical analysis was performed. As a result, genes and diseases associated with specific biological resources have been identified. In this study, through a comprehensive analysis of species, genes, and disease information using text mining, we were able to determine the utilization value of biological resources. This study will help the freshwater biodiversity researchers by adding a function for utilization analysis in the utilization information system of the platform in the future.

Keywords

utilization, freshwater biodiversity, platform, text mining

Presenting author

Ji-Hae Lee

Acknowledgements