

Metadata Management and Harvesting System in Smart Open Data As a Service

Minh Chau Nguyen*, HeeSun Won*, Siwoon Son*, Yerin Woo**, Hyeonjae Jeong**

*CybreBrain Research Section, Electronics and Telecommunication Research Institute, Daejeon, South Korea

** University of Science and Technology, Daejeon, South Korea

chau@etri.re.kr, hswon@etri.re.kr, siwoon.son@etri.re.kr, yrwoo@etri.re.kr, presenti94@etri.re.kr

Abstract—These days, with the available data platforms, the enormous growing information including data and metadata in big data and artificial intelligence fields can be stored, organized, and processed to solve different issues in different domains. However, there are still many limitations in the current platforms. Among these drawbacks, data and metadata exchanging interfaces between platforms, an essential process of information searching development, needs to be enhanced significantly. The main reason is that the emergence of too many platforms which follow totally different standards and structures has been making the process become more and more complicated and costly. Therefore, we propose here a system related to the platform management, in particular to information harvesting management to let users obtain easily the desired metadata and data content collected from different systems. Moreover, it helps users manage effectively the harvesting process.

Keyword—Database, Datahub, Harvesting, Metadata Management, Open Data



Minh Chau Nguyen received B.S. degree (2009) in Computer Science from the University of Sciences, Ho Chi Minh, Vietnam and M.S. degree (2013) in Computer Science from the Korea Advanced Institute of Science and Technology (KAIST). He is currently a senior researcher of the CybreBrain Research Section at Electronics and Telecommunications Research Institute (ETRI). His research interests include big data management, architecture, and distributed.



Heesun Won received B.S.(1990) degree in Computer Science from Yonsei University and the M.S.(1992), Ph.D. (2016) degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST). From 1992 to 1999, she was a researcher in Korean Broadcasting System. She is currently a principal researcher of CybreBrain Research Section at Electronics and Telecommunications Research Institute (ETRI). Her research interests include intelligent data hub, dataset curation, and *aaS platform.



Siwoon Son received B.S.(2014), M.S. (2016), and Ph.D. (2020) degrees in Computer Science from Kangwon National University. He is currently a postdoctoral researcher of the CybreBrain Research Section at Electronics and Telecommunications Research Institute (ETRI). His research interests include data mining, big data analysis and management, distributed storage systems, distributed processing systems, and distributed database systems.



Yerin Woo received B.S. degree (2019) in Engineering Information&Computer Engineering Computer Engineering from the Hongik University. She is currently a researcher of the CybreBrain Research Section at Electronics and Telecommunications Research Institute (ETRI). Her research interests include big data management, intelligent data hub and dataset curation.



Hyeonjae Jeong received B.S. degree (2020) in Computer Engineering from the Inje University. He is currently a researcher of the CybreBrain Research Section at Electronics and Telecommunications Research Institute (ETRI). His research interests include big data management, intelligent data hub and identity access management.