A web application ”Starrydata” for collecting and sharing plot data on published papers

Masaya Kumagai∗†1,2, Yukari Katsura‡3,4, Mitsunori Kaneshige5, Takushi Kodani3,4, Hideyasu Ouchi1,3, Sakiko Gunji4, Yuki Ando4, Yoji Imai1,4, Kaoru Kimura3, and Koji Tsuda1,3,4

1RIKEN Center for Advanced Intelligence Project – 15th Floor, Nihonbashi 1-chome Mitsui Building, Chuo-ku, Tokyo, Japan
2SAKURA Internet Inc. – 35th floor, Grand Front Osaka Tower A, 4-20 Ofukacho, Kita-ku, Osaka, Osaka, Japan
3The University of Tokyo – 5-1-5 Kashiwanoha, Kashiwa, Chiba, Japan
4National Institute for Materials Science (NIMS) – 1-2-1 Sengen, Tsukuba, Ibaraki, Japan
5X-Ability Co., Ltd. – 3rd floor, Ishiwata Building, 4-1-5 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

New materials have been discovered and designed by conventional concept-driven approaches based on the past experiences and knowledge. Lately, however, new approaches are expected because the complexity of new material discovery and design has a serious problem. In particular, Materials Informatics (MI), a data-driven approach to find new regularity and knowledge from an enormous data, is being rapidly developed.

In MI, first-principles calculations are used in many cases because data can be easily collected. In addition, as the enormous data have been shared on the Internet, MI has greatly contributed to the promotion of new material discovery and design. However, the values calculated under the ideal conditions by first-principles calculations are often different from the values obtained in experiments. Therefore, MI using experimental data is required, but hardly reported. Experimental data for MI are also hardly reported.

In this study, we have constructed an efficient web application named Starrydata, to collect and share experimental data on published papers. We extracted the experimental data as plotted, and comprehensively collected the sample compositions from the text. From 2,337 samples, experimental physical properties (electrical conductivity, thermal conductivity, Seebeck coefficient, etc.) have been obtained so far. We will introduce the details of this system and data, and application of this data for MI.

Keywords: thermoelectric materials, materials informatics, database, machine learning