

## SOLDER ALLOYS DATA MINING FOR MATERIALS RESEARCH

*M. Khvostova*<sup>1,\*</sup>, *A. Abramov*<sup>1</sup>, *A. Artamonov*<sup>1</sup>

<sup>1</sup> National Research Nuclear University MEPhI, Moscow, Russia

The study proposes a method for the automated processing of full-text PDF articles in the field of materials science that leverages visual, tabular, and textual data extraction techniques developed by the authors to obtain standardized representations of a selection of objects, thereby increasing the efficiency of technical data utilization. An integrative software tool built upon the method is discussed. The metrics of precision (63–99%) and recall (90–99%) calculated to assess tool performance prove the applicability of the solution proposed for domain-specific data aggregation, unification, search and management, and the potential for usage in science analytics systems.

Предложен метод автоматизированной обработки полнотекстовых статей по материаловедению в формате PDF, который сочетает техники извлечения изобразительных, табличных и текстовых данных, разработанные авторами, и направлен на получение стандартизированных представлений объектов из установленного перечня для повышения эффективности использования технической информации. Описано программное средство, основанное на данном методе. Рассчитанные метрики точности (63–99 %) и полноты (90–99 %) обработки публикаций с помощью программного средства подтверждают применимость предложенного решения для объединения, унификации, поиска и управления тематической информацией, а также возможность интеграции в системы научной аналитики.

PACS: 89.20.Ff

---

\* E-mail: mhvostova@kaf65.ru