

## Automatic Collection of Critical Temperature Related Expressions from Tables in Superconductors Development Publications

Thaer M. Dieb, Akira Suzuki, Luca Foppiano and Masashi Ishii

MaDIS, National Institute for Materials Science

E-mail: MOUSTAFADIEB.Thaer@nims.go.jp

Automatic construction of materials databases can support several applications in materials science domain. In this report, we propose an automatic extraction system to collect superconducting materials [1] and their critical temperature ( $T_c$ ) from tables in related publications using a rule-based approach and a chemical data extractor [2]. Tables provide condensed, rich source of structured information that is used to summarize experiment results.

A corpus of 9 documents in XML (JATS format) was collected using keyword search of superconductor/superconductivity, critical temperature, and filtered with the condition that table (defined by XML tag) contain  $T_c$ . Our table parser is applied to extract materials name, critical temperature expressions and values from tables. In many cases, only doping ratio is presented in the table, and not the sample name. In that case, we apply automatic linker to match the doping ratio with the sample name from the related table caption. Figure 1 illustrates the outline of the proposed system.

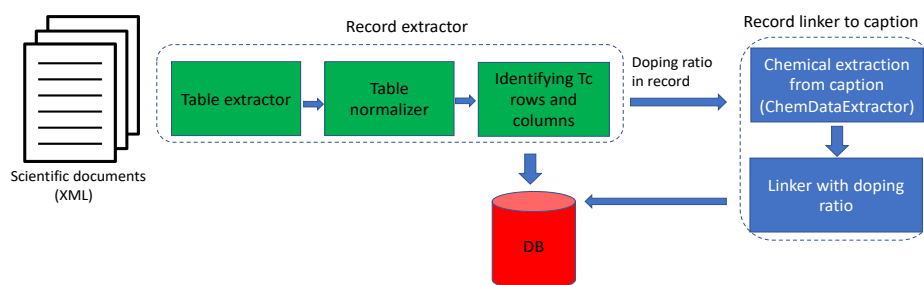


Figure 1: Extracting critical temperature from tables in superconductor related publications

For a better analysis, we evaluated the system from 2 perspectives. First entities extraction and then linking. Table 1 shows the system performance on entities extraction and abilities to link entities when necessary to the their samples from the caption.

	Precision	Recall	F score
Extraction	95.7%	100%	97.8%
Linking	48.6%	50.7%	49.6%

Table 1: System performance

In summary, the system demonstrated good recall for different types of tables; however, linking entities to their reference from caption is still a challenging task. In the future, we plan to continue developing additional rules to match entities in the tables to their original samples.

### References

- 1- Foppiano, L. & M. Dieb, T & Suzuki, A. & Ishii, M. Proposal of Automatic Extraction Framework of Superconductors related Information from Scientific literature. IEICE, pp:1-5 (2019)
2. Swain, M. & Cole, J. M. ChemDataExtractor: A toolkit for automated extraction of chemical Information from the scientific literature. Chem. Inf. Model 56, 1894–1904 (2016).