

## プラズマ CVD 技術文献のテキストマイニングを用いた 単語のインパクトの解析

### Analysis of Impact of Words using Text Mining of Papers regarding Plasma CVD

九大シス情<sup>1</sup>, 自然科学研究機構<sup>2</sup> °古閑 一憲<sup>1,2</sup>, 奥村 賢直<sup>1</sup>, 鎌滝 晋礼<sup>1</sup>,  
板垣 奈穂<sup>1</sup>, 白谷 正治<sup>1</sup>, 谷口 雄太<sup>1</sup>, 池田 大輔<sup>1</sup>

Kyushu Univ.<sup>1</sup>, NINS<sup>2</sup>, °Kazunori Koga<sup>1,2</sup>, Takamasa Okumura<sup>1</sup>, Kunihiro Kamataki<sup>1</sup>,  
Naho Itagaki<sup>1</sup>, Masaharu Shiratani<sup>1</sup>, Yuta Taniguchi<sup>1</sup>, and Daisuke Ikeda<sup>1</sup>

E-mail: koga@ed.kyushu-u.ac.jp

研究結果の考察や研究テーマの創出において、文献調査は重要である。多くの文献に当たること  
で知識の涵養と新しい概念の創発の助けとなる。しかしながら一人の研究者が、読み込むこと  
ができる文献の数には限りがあるため、情報の取り込み量は制限される。本研究では、プラズマ  
化学気相堆積(CVD)法文献のテキストマイニングから文献内単語の重要度を評価し、このダイナ  
ミクス解析とプラズマ技術者の協働により研究テーマの創発を捉えることを目的としている。こ  
こでは、文献に現れる単語のインパクトを示す指標を定義し、これを用いて単語のインパクトの  
時間変化を可視化することに成功した。

用いた文献情報は Web of Science で“plasma and chemical and vapor and deposition”をキーワード  
として検索して得られる 2000 年から 2018 年の文献情報の内、アブストラクトの文章を対象とし  
た。テキストマイニングと分析には、Python をベースとして Pandas, Numpy (データ処理ライブラ  
リ)、ScikitLearn (分析ライブラリ)、Holoviews, Seaborn (可視化ライブラリ) を用いた。文献に現  
れる単語のインパクトを評価するため、下記の正規化インパクト指数  $h(w,y)$  を定義した。

$$h(w,y) = \frac{f(w,y)}{\frac{p(y)}{\sum_{t=2000}^{t=2018} f(w,t)}}$$

ここで、 $w$  は単語、 $y$  は出版年、 $f(w,y)$  は  $y$  年に現れた単語  $w$  の頻出数、 $p(y)$  は  $y$  年に出版された文  
献数を示す。つまり、1 年あたりに出版された文献 1 本あたりに現れる単語の頻出数を期間全体  
での総頻出数で正規化したものである。得られ  
た  $h(w,y)$  について 2000 年を基準とした時間変  
化  $h(w,y)-h(w,2000)$  を可視化したものを図 1 に  
示す。この結果はプラズマ CVD 技術関連単語  
の注目の度合い (文献内で取りあつかわれた  
数) の時間変化を示すものである。講演では、  
論文数の時系列推移などと比較しつつ、単語の  
トレンドを議論する。

謝辞：本研究は、Clarivate Analytics の協力の下  
実施された。

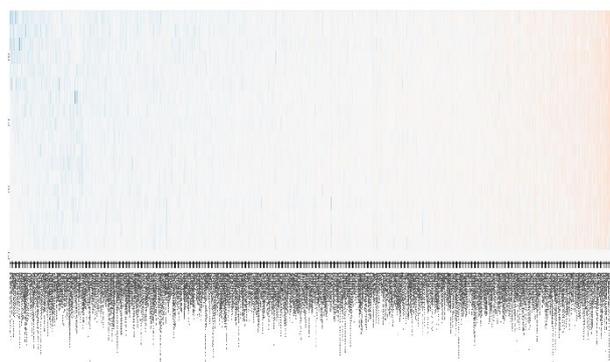


図 1. プラズマ CVD 関連文献に現れる単語  
の持つインパクトのヒートマップ。