

テキストマイニングを用いた 論文・社内報告書の分析事例

用語抽出ツールを用いた
取り組み内容のご紹介

株式会社村田製作所
長島 和輝





1. ムラタのご紹介
2. 用語抽出ツールを用いた取り組み内容のご紹介
3. 取り組みを行った結果
4. 今後の予定

1. ムラタのご紹介 ～ムラタの概要～



村田製作所は、最先端の技術、部品を創出する総合電子部品メーカーです。Innovator in Electronicsをスローガンに掲げ、豊かな社会の実現をめざします。

ムラタの強み

- 最先端の材料を研究開発
- 広範囲な製品ラインナップ
- グローバルな生産、販売ネットワーク

ムラタのプロフィール

- 創業： 1944年
- 売上高： 1兆6千301億9千3百万円
- 企業数： 89社（国内29社、海外60社）
- 従業員数： 75,184名（国内31,775名、海外43,409名）

※売上高は、2021年3月期決算。
※従業員数は2021年3月31日時点のものです。
※グループ企業数は2021年3月31日時点のものです。
※村田製作所はグループ企業数に含まれておりません。

1. ムラタのご紹介 ～市場とアプリケーション～

ムラタの技術が、 エレクトロニクスの 可能性を押し広げます

- スマートフォンやパソコン、AV機器、家電製品などのエレクトロニクスの中心分野から
- 自動車やエネルギー、ヘルスケアなどの領域へも

暮らしに安心、安全、便利を・・・

通信機器



- 基地局
- 固定無線アクセス機器
- G-PON
- ルータ・スイッチ

モビリティ



- 電気自動車
- インフォテインメント
- ADAS
- 安全システム
- 自動二輪車

エンタープライズシステム



- サーバ
- データセンタ
- ストレージシステム

産業用機器



- ファクトリーオートメーション
- セキュリティ・ビルオートメーション
- スマートグリッド

ヘルスケア・メディカル



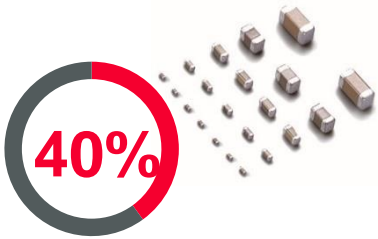
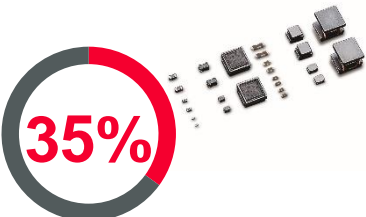
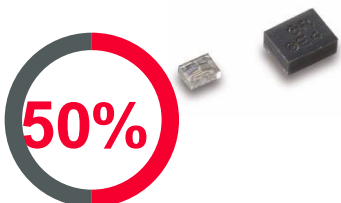
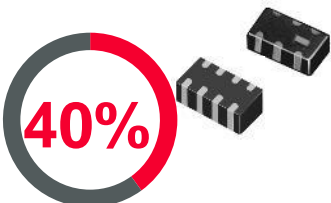

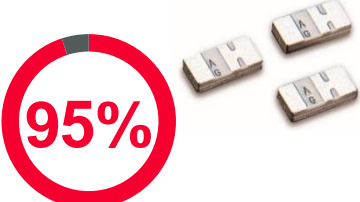
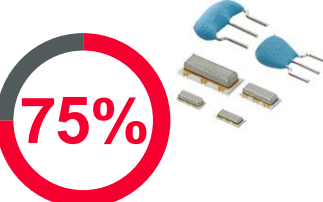
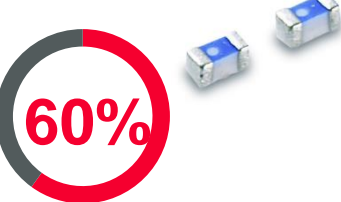
- パーソナルヘルスケア機器
- 医療用機器

パーソナルエレクトロニクス



- スマートフォン
- ウェアラブル・ヒアラブル
- PC・タブレット
- 家電製品

1. ムラタのご紹介 ～主要製品の世界シェア*～

 <p>40%</p>	 <p>35%</p>	 <p>50%</p>	 <p>40%</p>
チップ積層セラミックコンデンサ	EMI除去フィルタ (EMIFIL®)	表面波フィルタ	多層LCフィルタ (フィルタ、バラン、カプラなど)
 <p>40%</p>	 <p>95%</p>	 <p>75%</p>	 <p>60%</p>
マイクロ 배터리 「酸化銀電池」	ショックセンサ	セラミック発振子	高周波インダクタ

*主要製品の世界シェアは当社推定値です。また市場や用途により異なります。

1. ムラタのご紹介 ～ムラタの技術～

材料技術						
	材料設計	材料プロセス				
前工程技術						
	積層	印刷	焼成	薄膜微細加工	表面処理	精密加工
商品設計技術						
	高周波設計	デバイス設計	組み込み	高信頼性設計	回路設計	シミュレーション
後工程技術						
	パッケージング	計測	自動化設備	IE		
分析・評価技術						
	材料分析	故障解析				

1. ムラタのご紹介 ～ムラタ内の機械学習活用～

✓ 過去

- ◆ 一部の人が個々の業務で先行的に活用している
- ◆ 業務効率化に機械学習を活用したい人はいるが、ハードルが高い



✓ 現在

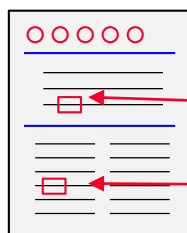
- ◆ 機械学習の体系的な理解
- ◆ 機械学習の業務活用

を目的とした社内教育講座を、2019年よりNTTデータ数理システム様に
依頼して開催中

長島・藤田は二期生（2020年度）として受講

2. 用語抽出ツールを用いた取り組み内容のご紹介 ～背景、課題と目的～

- ✓ 背景：新規プロセス技術の材料検討には多くの知見が必要であり、様々な専門知識を持った人が力を合わせて調査を行っている
 - ✓ 課題：近年はより複雑な化学合成となっているため、文献調査として読み解くのに膨大な時間を要している
 - ✓ 目的：事前収集した金属/セラミックス/樹脂材料の化学合成に関する文献からの調合表（「生成物」「原材料」「調合量」「調合条件」）の用語抽出ツールの構築
- 第1ステップ：文献からの「生成物」「原材料」の用語抽出ツールの構築
- 第2ステップ：文献からの「調合量」「調合条件」の用語抽出ツールの構築



文献

(例)

生成物：銀

原材料：硝酸銀、純水、エタノール、ポリビニルピロリドン

2. 用語抽出ツールを用いた取り組み内容のご紹介 ～使用ツール～

- ✓ テキストマイニングスタジオ (TMS) : NTTデータ数理システム様開発

The screenshot shows the TextMiningStudio (TMS) interface. The main window displays a list of words and their frequencies, along with a table of extracted groups. The left sidebar shows various analysis tools like '単語頻度解析' and 'グループビンギ'. The right sidebar shows a '結果リスト' (Result List) with a tree view of the extracted groups.

単語	頻度
particle	650
size	496
nanoparticle	451
temperature	447
use	412
show	394
have	371
solution	343
figure	321
sample	312
high	295
powder	284
result	276
obtain	273
reaction	252
method	248
metal	247

グループ名	頻度	グループ
1_Ag	71	グループ
2_Au	41	グループ
3_Co	18	グループ
4_Mn	28	グループ
5_Ni	142	グループ
6_Pd	28	グループ
7_Fe	14	グループ
8_Ru	10	グループ
9_FeCo	5	グループ
10_Co7Fe3	2	グループ
11_Co-Cu	32	グループ
12_CoNi	12	グループ
13_CoPt	1	グループ
14_NiO	0	グループ

- ・ 自然言語処理のための汎用ソフトウェア
- ・ コーディングが不要

2. 用語抽出ツールを用いた取り組み内容のご紹介 ～取り組みのステップ～

前処理

- ・ 文献10-20本程度の読み込み → グルーピングルールを作成
- ・ 文献の、TMSで読込可能な形式 (txtファイル) への変換

表計算マクロ

グルーピング

- ・ 文献 (txtファイル) とグルーピングルールを用いた用語抽出

TMS

後処理

- ・ TMSからの出力ファイルの整理

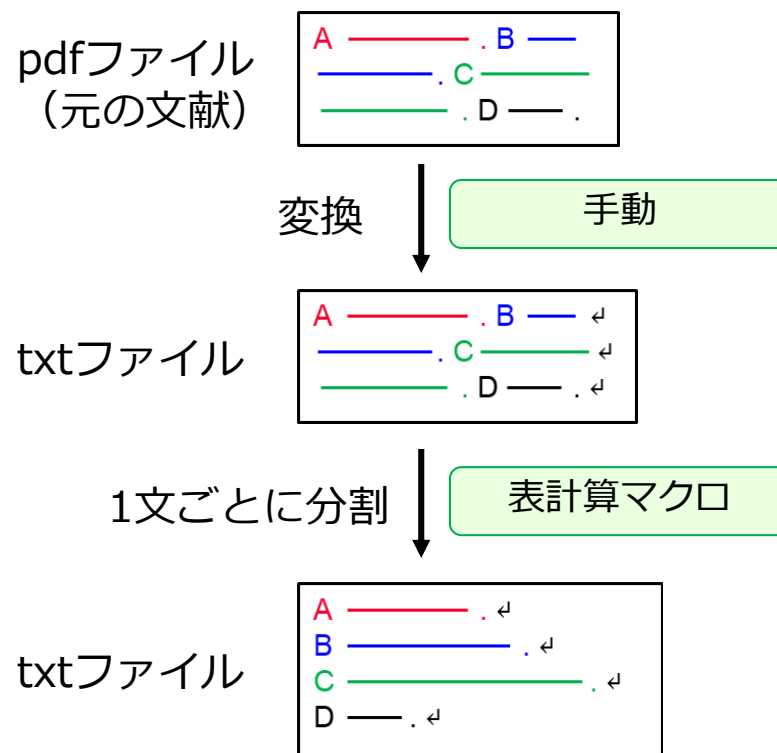
表計算マクロ

2. 用語抽出ツールを用いた取り組み内容のご紹介 ～前処理～

✓ グループインクルール (抽出したい用語の候補) の作成

	A	B
1	#dataname 210811_グルーピング1	
2		
3	Ag->particle	1_Ag
4	Ag->nanoparticle	1_Ag
5	Ag->powder	1_Ag
6	nano->Ag	1_Ag
7	Ag->nanocrystals	1_Ag
8	Ag->nanowire	1_Ag
9	Ag->NPs	1_Ag
10	Ag->nanoballs	1_Ag
11	silver->particle	1_Ag
12	silver->nanoparticle	1_Ag
13	silver->powder	1_Ag
14	nano->silver	1_Ag
15	silver->nanocrystals	1_Ag
16	silver->nanowire	1_Ag
17	silver->NPs	1_Ag
18	silver->nanoballs	1_Ag
19	Au->particle	2_Au
20	Au->nanoparticle	2_Au
21	Au->powder	2_Au
22	nano->Au	2_Au
23	Au->nanocrystals	2_Au

✓ 文献をTMSで読込可能な形式に変換



2. 用語抽出ツールを用いた取り組み内容のご紹介 ～グルーピング～

グルーピングルール内の用語のうち、文献内に使用されているものが抽出される

行番号

その行に記述されている「生成物」

文献番号

ファイルID	行ID	文章ID	material	composite	原文	material_key
1	55	-1	無	-1	The dye could	<input type="checkbox"/>
1	56	-1	1_H2O	無	None of these	<input type="checkbox"/>
1	57	-1	無	1_Ag	Since more tha	<input type="checkbox"/>
1	58	-1	無	無	This leads us t	<input type="checkbox"/>
1	59	-1	無	無	The former expl	<input type="checkbox"/>
1	60	-1	無	無	Results similar	<input type="checkbox"/>
1	61	-1	無	無	The absorption	<input type="checkbox"/>
1	62	-1	無	1_Ag	The amount of	<input type="checkbox"/>
1	63	-1	1_H2O	2_Au	Thus 90 % of	<input type="checkbox"/>
1	64	-1	無	無	As mentioned	<input type="checkbox"/>
1	65	-1	無	無	The spectrum	<input type="checkbox"/>
1	66	-1	5_NaBH4	無	The same spec	<input type="checkbox"/>
1	67	-1	無	無	Close comparis	<input type="checkbox"/>
1	68	-1	無	無	Even at every l	<input type="checkbox"/>
1	69	-1	無	無	Such a spectru	<input type="checkbox"/>
1	70	-1	無	無	Although the si	<input type="checkbox"/>
1	71	-1	無	無	Attempts to cb	<input type="checkbox"/>
1	72	-1	1_H2O	1_Ag	Note that altho	<input type="checkbox"/>

その行に記述されている「原材料」

2. 用語抽出ツールを用いた取り組み内容のご紹介 ～訴求ポイント～

- ✓ 本取り組みでは、「事前準備を技術者」「用語抽出作業を技術補助職」として全体効率化を考えており、その側面で、2項目に分けて記載する

項目		従来手法 (読み込み)	今回の手法 (用語抽出ツール)	今回の手法の特徴
事前準備 (技術者)	内容	なし	20件程度の文献の読み込み、グルーピングルールの作成	✗ 事前準備が必要
	所要時間	—	3時間程度	
	必要な知識	なし	自然言語処理の専門知識、日本語/英語の読解力	✗ 事前準備には専門知識が必要
用語抽出 作業 (技術 補助職)	正確さ (F1 score)	0~1 (作業者に依存する)	0.93	✓ 正確さは、高い水準で安定している
	工数	10分程度 / 1文献	2分以下 / 1文献	✓ 工数が80%以上削減できる
	必要な知識	日本語/英語の読解力	なし	✓ 用語抽出ツール構築後は専門知識が不要

3. 取り組みを行った結果

- ✓ TMSはコーディングが不要であるため、自然言語処理を試してみるのに良いツールである
- ✓ 自然言語は記述に関して統一ルールがなく、表記ゆれ、誤字、難解な表現などを取り扱う上で困難なポイントは多いが、試行錯誤を繰り返して目標に合致した動作をさせたので、成果を出すことができた

4. 今後の予定

- ✓ 本取り組みは2021年1月より継続してきたが、ようやく成果が出てきた段階であり、今後も以下の活動に継続して取り組む
 - ・ 今回の取り組みで明確になった課題
 - ◆ 原材料以外の試薬（洗浄液など）を原材料として抽出しない
 - ◆ 不定比化合物に対応する
 - ・ 最終的な目標である「調合量」「調合条件」の用語抽出ツールの作成
 - ・ 今回の用語抽出ツールの他の活動への展開
- ✓ 本事例を周囲に報告することで、具体的イメージを掴むことができたようで興味を持ってくれる方が増えた
- ✓ 機械学習の社内教育講座は今後も継続的に新規メンバーに対して実施されている
- ✓ 取り組み内容の近いメンバーとは進捗・課題を共有しながら、今後とも、機械学習の使用領域を広げていく必要がある



ご清聴ありがとうございました