

# DDA マニュアル ③

## ～ファイルとデータの操作・機械学習編～

Clarivate

Jan 26 /2023

Derwent Data Analyzer v.12

# Derwent Data Analyzer (DDA)

## ③ファイルとデータの操作・機械学習編

様々なデータソースからデータを取り込む

- Web of Science の文献データを取り込む
- Excel形式のデータを取り込む
- データのさらなる整理方法

データの追加：複数ファイルの統合

- 後からフィールドを追加する
- 複数のファイルを結合する

機械学習機能を利用して  
類似特許を探す

- 特定の特許から近い内容のレコードを探す
- 文章入力から、近い内容のレコードを探す

機械学習機能を利用して  
自動分類を行う

- 教師データの作成
- 自動分類

様々な処理：

- Excelへのデータ出力
- シソーラスの活用

様々なデータソースからデータを取り込む

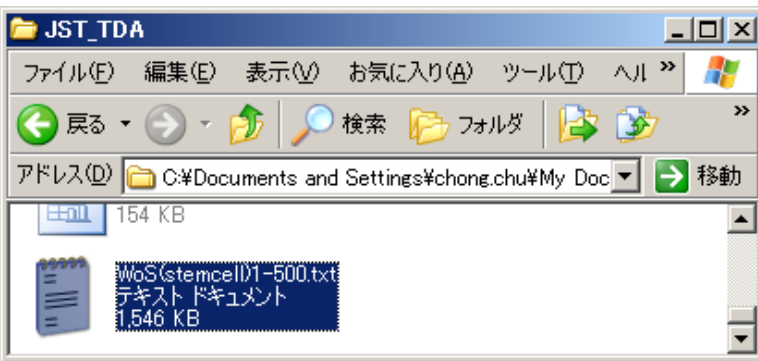
# Web of Scienceの学術文献データを取り込む

Web of Science、INSPEC等、同じ方法で行うことができます。  
エクスポートの方法についての詳細は、Web of Scienceのマニュアルを参照してください。

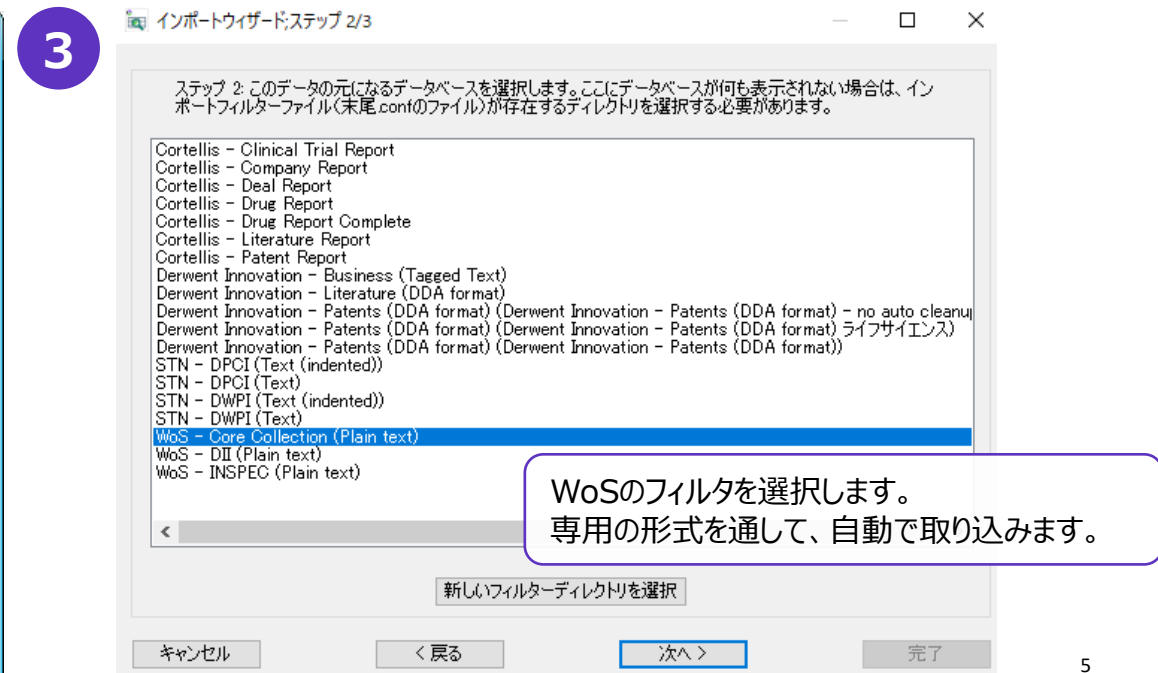
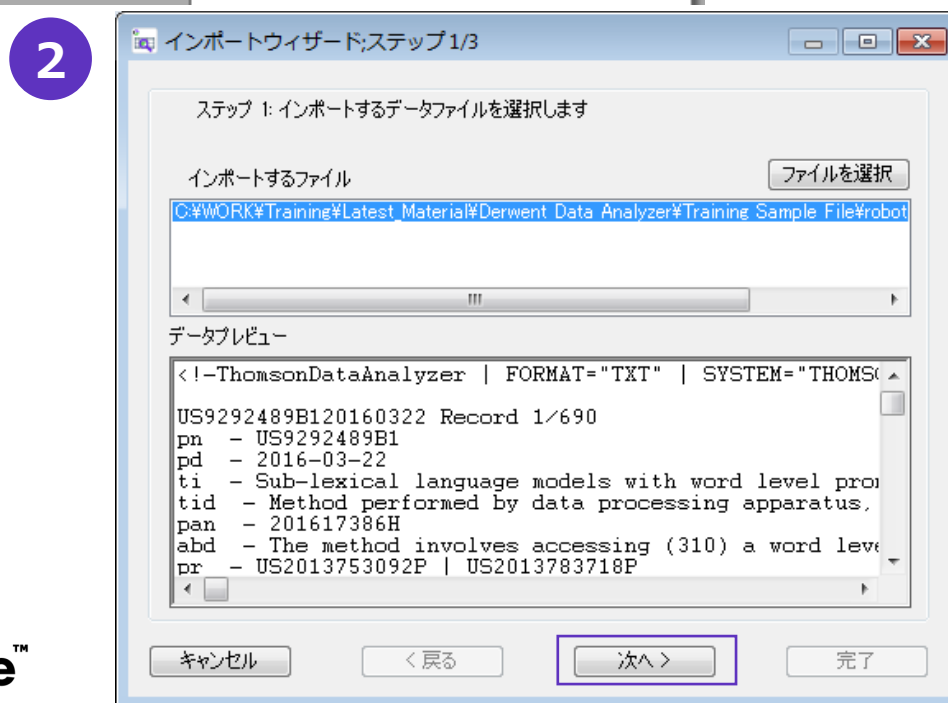
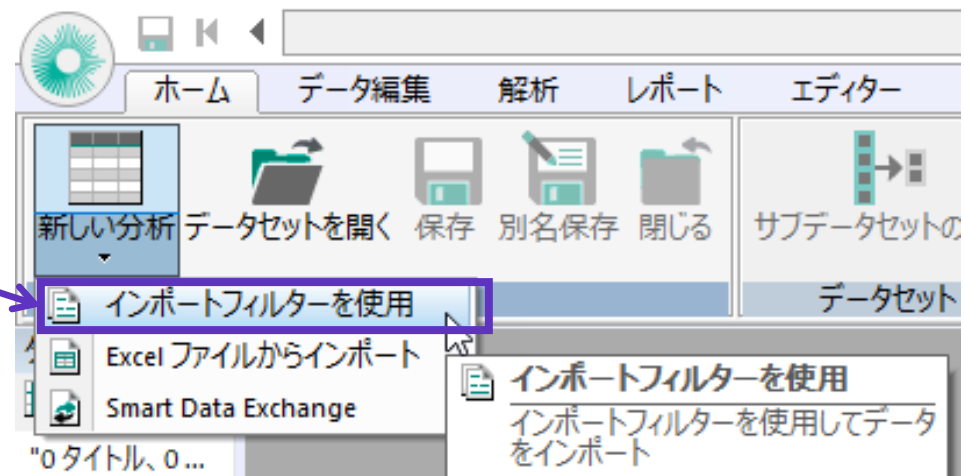
## 1 Web of Science から「テキストで保存」形式を指定してダウンロードします

The screenshot shows the Web of Science search results page. At the top, there are two search results. The first result is titled "Galicia and the Bay of Biscay: Lysosomal responses" and the second is "Mangrove Bacterial Diversity and the Impact of Oil Contamination Revealed by Pyrosequencing: Bacterial Proxies for Oil Pollution". Below the search results, there is a section for "レコードの出力" (Record Output) with three steps: Step 1 (Select records), Step 2 (Select fields), and Step 3 (Select export format). In Step 3, the "テキストで保存" (Save as text) option is selected. The "保存先" (Save to) dropdown menu is open, showing options like "文献管理ソフト向けに保存" (Save for literature management software), "BioTeXで保存" (Save with BioTeX), "HTMLで保存" (Save as HTML), "テキストで保存" (Save as text), "タブ区切りで保存(Win)" (Save as tab-delimited (Win)), and "タブ区切りで保存(Mac)" (Save as tab-delimited (Mac)). The "テキストで保存" option is highlighted.

## 2 ファイルを適切な場所に保存



# Web of Scienceの学術文献データを取り込む



# Excel形式のデータをDDAに取り込む

下の例のように、最初の行に各列名があり、2行目以降に各レコードの該当のフィールド情報が入った形式ファイルであれば利用可能です。

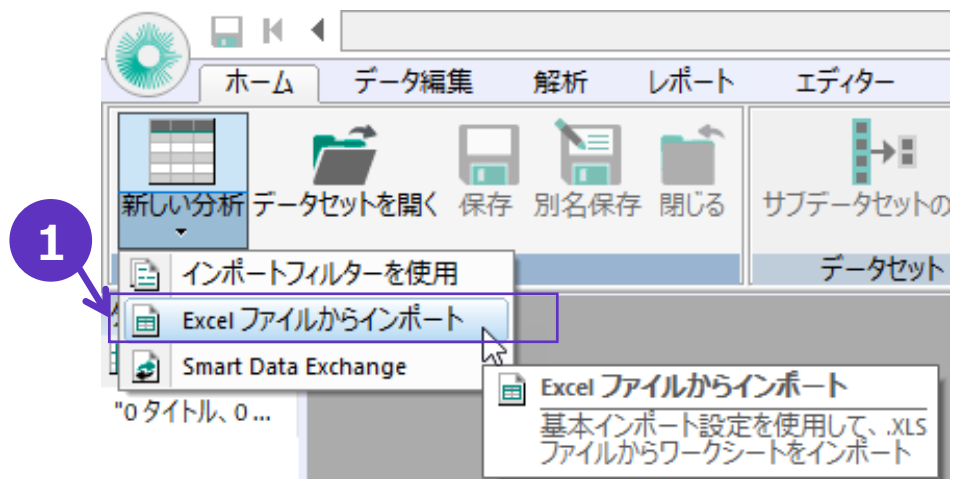
様々なデータベースでダウンロードした情報をDDAにインポートすることが可能です。

## 【例：Cortellis からダウンロードしたファイル】

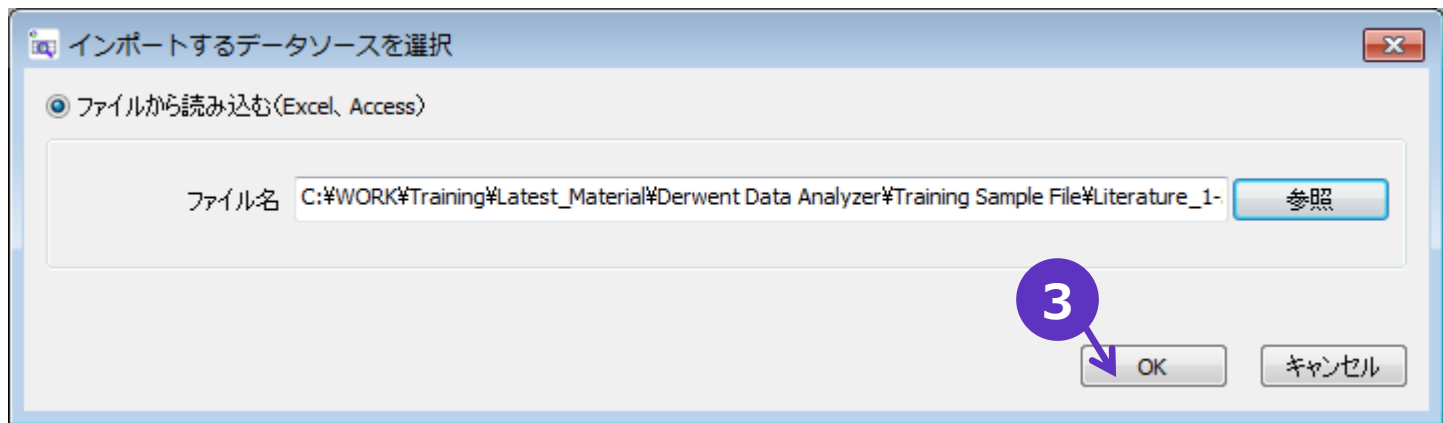
Patent Number	Publication Date	Company (assignee)	Title	Therapy Areas	Actions	Technologies	Drug	Classification
US-07022724	04-Apr-2006	Merck Sharp & Dohme Ltd	New isoxazole triazoloindane derivatives are inverse agonists of the GABA <sub>A</sub> receptor - useful for the treatment of diminished					
WO-2006032273	30-Mar-2006	H Lundbeck A/S	New 2-acylaminothiazole derivatives are adenosine A <sub>2A</sub> receptor modulators - useful for the treatment of neurological and psychiatric disorders.	Alzheimer's disease; Bleeding; Brain hemorrhage; Brain injury; Cerebral hypoxia; Cerebrovascular ischemia; Epilepsy; Huntingtons chorea; Hypoxia; Ischemia; Major depressive disorder; Parkinsons disease; Psychotic disorder	Adenosine A2a receptor modulator			Product
WO-2006034296	30-Mar-2006	CoMentis Inc; Oklahoma Medical Research Foundation; Zapaq Inc; University of Illinois	New amino containing compounds are inhibitors of memapsin-2 $\beta$ -secretase activity - useful for the treatment of Alzheimer's disease.	Alzheimer's disease	Beta secretase inhibitor; Neuroprotectant			Product
WO-2006034089	30-Mar-2006	Targacept Inc	New azaspiroalkene and azaspiroalkane compounds are nicotinic cholinergic receptor modulators - useful for the treatment of CNS disorders such as pre-senile dementia, Alzheimer's disease, HIV dementia and Parkinson's disease.	Alzheimer's disease; Attention deficit hyperactivity disorder; Central nervous system disease; Cognitive disorder; HIV associated dementia; Huntingtons chorea; Major depressive disorder; Mania; Niemann Pick disease; Parkinsons disease; Schizophrenia; Alzheimer's disease; Atherosclerosis; Crohns disease; Dermatological disease; Hypercholesterolemia; Hyperglycemia; Hyperlipidemia; Irritable bowel syndrome; Lipid metabolism disorder; Non-insulin dependent diabetes; Obesity; Restenosis;	Antipsychotic; CNS modulator; Integrin alpha-7/beta-2 modulator; Nicotinic acetylcholine receptor modulator; Nootropic agent			Product
WO-2006033891	30-Mar-2006	Merck & Co Inc	New benzisoxazole derivatives are PPAR $\alpha$ agonists - useful for the treatment of lipid disorders and for reduction of the risk of development of atherosclerosis.		PPAR agonist; PPAR alpha agonist			Product
WO-2006034277	30-Mar-2006	CoMentis Inc; Zapaq Inc; University of Illinois; Oklahoma Medical Research Foundation	New bicyclic compounds are inhibitors of memapsin-2 $\beta$ -secretase activity - useful for the treatment of Alzheimer's	Alzheimer's disease	Beta secretase inhibitor; Neuroprotectant			Product
WO-2006033318	30-Mar-2006	Astellas Pharma Inc	New cyclic amine derivatives are NMDA antagonists - useful for treating and preventing senile dementia.	Alzheimer's disease; Cerebrovascular ischemia; Pain; Parkinsons disease; Senile dementia	NMDA receptor antagonist			Product
WO-2006034317	30-Mar-2006	Glaxo Wellcome plc; SmithKline Beecham Corp	New indole carboxamide compounds are IKK2 inhibitors - useful for the treatment of e.g. rheumatoid arthritis, inflammatory bowel disease, asthma, psoriasis, systemic lupus erythematosus, multiple sclerosis, Alzheimer's disease, cancer and viral infection.	Acquired Immune Deficiency Syndrome; Alzheimer's disease; Arthritis; Asthma; Autoimmune disease; Cancer; Chronic obstructive pulmonary disease; Dermatological disease; Inflammatory bowel disease; Multiple sclerosis; Neurodegenerative disease; Psoriasis; Respiratory disease; Rheumatoid arthritis; Systemic lupus	I-kappa B kinase beta inhibitor			Product

# Excel形式のデータをDDAに取り込む

## 例：Excelデータをインポートする



2 インポートするExcelファイルを選択します。



# Excel形式のデータをDDAに取り込む 例：Excelデータをインポートする

3

Excel中の各フィールドのインポート方法を選択します。

## 注) 先頭行の指定

先頭行の各フィールド名が記載されている行を指定します。  
正しく指定されると、下段の表に読み込まれた各フィールド名が表示されますので、確認してください。  
各フィールドをクリックすると、読み込まれた次行移行の各データを確認することができます。

データベースからクイックインポート

フィールド名を含む行 1 リフレッシュ ☒ 常にこの行を使用

フィールド名	フィールドタイプ	区切り文字
Abstract Classification	テキストを分割	
Advantage	単一値	
Annotation	単一値	
Application Date	単一値	

フィールド名: Abstract Classification  
フィールドタイプ: テキストを分割 区切り文字: デフォルトとして保存  
設定

フィールドプレビュー

Biologicals and Immunologicals  
Cardiovascular and Renal  
Oncologic, Endocrine and Metabolic

< 前のレコード 1 次のレコード >

☐ インポートフィルターをファイルに保存 参照

インポート キャンセル

## 注) Field Typeの変更

1つのフィールドに複数値が含まれる場合は、**Divide Text** を選択し、**Delimiter**で区切り文字を指定します。  
すると、区切られた値が縦に表示されます。  
正しく区切られていることを確認してください。

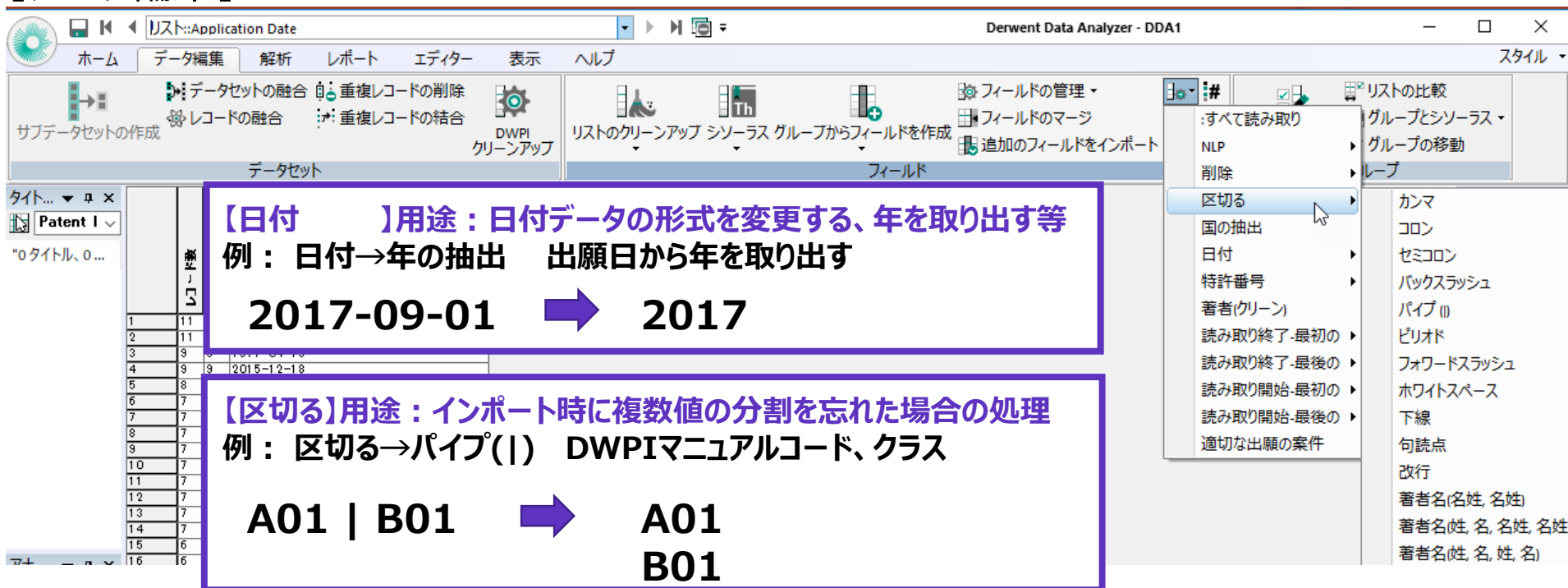
文章形式の場合は、**NLP** を指定します。

取り込んだ後からでも行えるので、必ずしもここで行う必要はありません。



# 外部ファイルを取り込んだ際に必要となる様々なデータ加工①

## 【データ編集】



**【日付】用途：日付データの形式を変更する、年を取り出す等**  
例：日付→年の抽出 出願日から年を取り出す

2017-09-01 ➡ 2017

**【区切る】用途：インポート時に複数値の分割を忘れた場合の処理**  
例：区切る→パイプ(|) DWPIマニュアルコード、クラス

A01 | B01 ➡ A01  
B01

**【読み取り開始】用途：必要な部分のみを抜き出す**  
例：読取り開始 - 最後の → comma(,) 住所など

Clarivate Tokyo, Japan ➡ Japan

## 外部ファイルを取り込んだ際に必要となる様々なデータ加工②

### 【メタタグ（フィールドの属性）設定】

例：country情報（国コード）をワールドマップの軸に設定したい場合

サマリーシート					
レコード数: 1,017					
フィールド	項目数	グループ数	カバー率 (%)	データタイプ	メタタグ
(フィルター)					
Company Name	1,017		100%		
Country	56		100%		
Field of Activity	73		45%		
Key Indications	579		57%		
Key Target-based Actions	582		57%		
Key Technologies	593		62%		
Number of Deals	152		100%		
Number of Drugs in Active Development	86		100%		
Number of Inactive Drugs	128		100%		
Number of Patents as Owner	334		100%		
Number of Patents as Third Party	86		100%		
Organization Type	18		62%		
Parent Company Name	528		99%		

1 属性を設定したいフィールドのメタタグエリアを右クリック

- リストの作成(L)
- リストのクリーンアップ...
- シソーラス...
- 検索と置換...
- 追加データ処理
- マイキーワードを抽出
- フィールド名の変更...
- フィールドのコピー...
- データタイプを設定(S)
- メタタグを設定(M)...
- フィールドの削除...
- 統計データを表示(V)...
- ズーム

2

メタタグの追加/削除

フィールド: Country

利用可能なメタタグ

- Category
- Child
- Child (Inactive)
- Classification
- Concept
- Concept General
- Concept Specific
- Country
- Date
- DDA Derived
- Document Characterization
- Document Source
- Document Type
- Incremental Import
- Language
- Long Text
- Organization
- Parent
- Person
- Record Title
- Short Text
- TDA Derived
- Training Field
- Unique ID
- Year

選択したメタタグ

Country

3 属性を選択して右側に移動させ、OK

OK キャンセル

マップの機能やレポートの機能は、メタタグによって、情報の種類を認識して動いています。

DI, WOS以外の外部からの情報は、メタタグがセットされないため、自分でセットする必要があります。以下を設定するとよいでしょう。

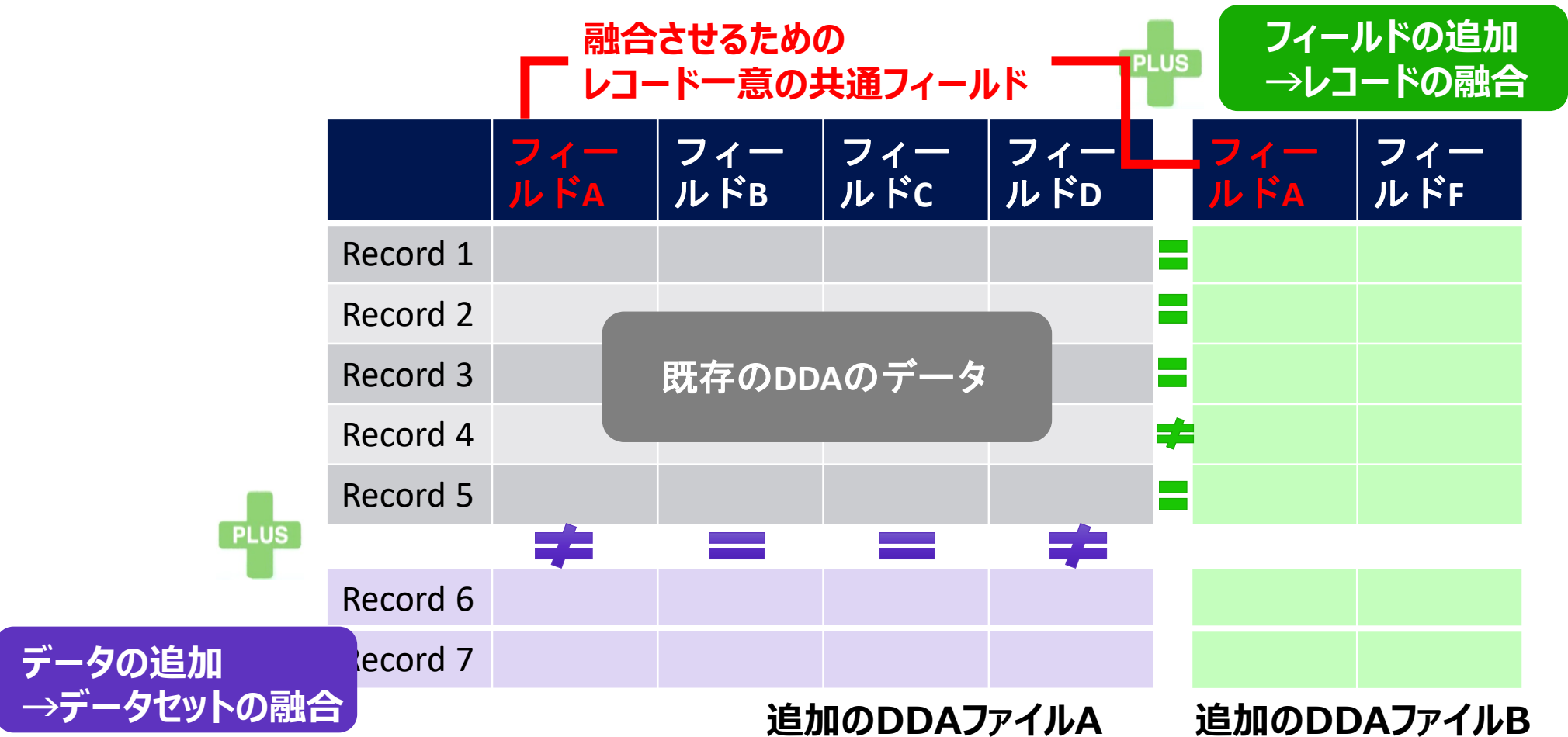
- Country
- Year
- Unique ID
- Person
- Organization

# データの追加：複数ファイルの統合

- ・後からデータを追加する
- ・後からフィールドを追加する

# データ追加の考え方

DDAでは、最初に取り込んでいなかったデータなども後から取り込み、付け足すことが可能です。



## 後からデータを追加する（データセットの融合）

データの追加を行います。同じフィールドを持っている必要はなく、異なるデータソースのデータも取り込めます。

例) Derwent Innovationの特許キーワードと文献キーワードを併せて分析する

例) 既存の分析ファイルに新しい年代のデータを追加する

例) 一度にダウンロードできず複数回に分けてダウンロードしたファイルを1つにまとめる

データの追加  
→データセットの融合

The screenshot shows the Derwent Data Analyzer - DDA4 software interface. The 'データ編集' (Data Edit) tab is active, and the 'データセットの融合' (Data Set Fusion) button is highlighted with a blue box and a purple arrow labeled '2'. Below the toolbar, the 'サマリーシート' (Summary Sheet) is displayed, showing a table with columns: フィールド (Field), 項目数 (Number of Items), グループ数 (Number of Groups), カバー率 (%) (Coverage (%)), データタイプ (Data Type), and メタタグ (Meta-tag). The table lists fields like 雑誌名 (Magazine Name), 著者 (Author), 出版年 (Publication Year), タイトル (Title), and UT, with their respective counts and coverage percentages. A purple box labeled '1' points to the '元のDDAのファイルと追加のデータファイルを両方開いておきます。' (Open both the original DDA file and the additional data file.) instruction. Below this, two file icons labeled 'DDA3' and 'DDA4' are shown, with 'DDA4' highlighted by a red box.

フィールド	項目数	グループ数	カバー率 (%)	データタイプ	メタタグ
(フィルター)					
雑誌名	45		100%		
著者	386		100%		
出版年	22		100%	番号	
タイトル	398		100%		
UT	398		100%		

## 後からデータを追加する（データセットの融合）

データセットを結合

フィールドの追加先: DDA4 **3** 最終的にマージする先のファイル

フィールドに追加するデータを選択: DDA3 **4** 追加するデータが含まれるファイル

利用可能フィールド: DDA3

UT  
出版年  
著者  
雑誌名

結合されたデータセット

- UT
  - [DDA4] -> UT
  - [DDA3] -> UT
- タイトル
  - [DDA4] -> タイトル
  - [DDA3] -> タイトル
- 出版年
  - [DDA4] -> 出版年
- 著者

**5** 左のフィールドからマージ対象のフィールドを選択して、右に入れる。  
最終ファイルのフィールドの状況を確認する。

フィールド名の変更  
フィールドの削除  
フィールドを復元

▼ 詳細設定

☒ グループを保持

☒ 同じ名前でグループを結合

※作成したあったグループなどは引き継がれます。

不確定グループタグの処理  
新リストにマークしない（確実性-高）

**6** OK キャンセル

## 後からフィールドを追加する（レコードの融合）

既存のDDAファイルにあるレコードについて、現在のフィールドにはない、新しいデータ項目（フィールド）を追加したいときに行います。追加するデータは、予め用意し、DDAのファイルとして用意しておく方法です。

例） Derwent InnovationのデータとCortellisの特許データを併せて分析する

例） 最初に取り込まなかったデータ項目を後から追加する

フィールドの追加  
→レコードの融合

The screenshot shows the Derwent Data Analyzer software interface. The title bar indicates the file is '追加フィールド.vpt'. The menu bar includes 'ホーム', 'データ編集', '解析', 'レポート', 'エディター', '表示', and 'ヘルプ'. The toolbar contains various icons for data manipulation, with a callout box highlighting the '既存のデータセットからレコードを融合' (Merge records from existing dataset) option under the 'データ編集' (Data Edit) menu. A second callout box points to the 'マスターファイル.vpt' and '追加フィールド.vpt' files in the taskbar, with the instruction: '元のDDAのファイル（マスターファイル）と追加のデータファイル（追加フィールド）を両方開いておきます。' (Open both the original DDA file (master file) and the additional data file (additional field)).

The main window displays a table with the following data:

フィールド	項目数	グループ数	カバー率 (%)	データタイプ	メタタグ
(フィルター)					
要約	884		98%		
住所	833		95%		
キーワード索引	741		82%		
UT	897		100%		

The table also shows a status bar at the top indicating 'レコード数: 897' (Record count: 897) and a '列表示' (Column display) button.

## 後からフィールドを追加する（レコードの融合）

既存のデータセットからレコードを融合

フィールドの追加先: マスターファイル.vpt **3** 最終的にマージする先のファイル

フィールドに追加するデータを選択: 追加フィールド.vpt **4** 追加するフィールドが含まれるファイル

このマッチングレコードのフィールドを選択: マスターデータセット および アクセサリーデータセット

アクセサリデータセット	必須	マスターデータセット	一致タイプ	マッチングルールセット
UT	完全一致	UT	文字列	文字列

+ マッチング基準を追加

利用可能フィールド: 追加フィールド.vpt

入力して検索...

要約

**6** 左のフィールドからマージ対象のフィールドを選択して、右に入れる。  
最終ファイルのフィールドの状況を確認する。

新規融合されたデータセット

- UT  
[マスターファイル.vpt] -> UT  
[追加フィールド.vpt] -> UT
- + ファイル
- + 出版年
- + 著者
- + 雑誌名
- キーワード索引  
[追加フィールド.vpt] -> キーワード索引
- 住所  
[追加フィールド.vpt] -> 住所

**5** 共通項フィールドは、文字列で完全一致と設定する

共通項フィールド名は異なっても可だが合わせておくと良い

出来上がりのファイルのフィールド状態の一覧

すべてのフィールドを展開 すべてのフィールドを折りたたむ

フィールド名の変更  
フィールドの削除  
フィールドを復元

詳細設定

**7** OK キャンセル



# レコード融合を利用すると、別々のデータベース情報を掛け合わせた分析も可能になります

Reset		DWPI Manual Codes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		# Records	56	19	18	14	13	11	97	96	96	96	94	92	86	86	86	82	81	80	80	74	70	62
Therapy Areas (Cleaned)	# Records	Show Values >= 1																						
		Cooccurrence # of Records																						
			Alzheimer's disease	Parkinson's disease	Cancer	Neurodegenerative disease	Inflammatory disease	Cerebrovascular ischemia	Major depressive disorder	Obesity	Pain	Cognitive disorder	Multiple sclerosis	Diabetes mellitus	Cardiovascular disease	Atherosclerosis	Rheumatoid arthritis	Anxiety disorder	Autoimmune disease	Schizophrenia	Huntington's chorea	Asthma	Dementia	Neurological disease
1	29	B14-J01 A4	287	71	95	57	75	47	27	35	29	47	47	39	43	51	49	27	44	26	31	42	38	29
2	28	B14-N1 6	277	118	96	79	71	96	43	40	48	40	56	42	49	44	42	45	41	45	58	39	41	42
3	24	B14-J01	238	99	76	85	55	60	60	56	55	49	36	47	40	30	29	49	28	56	42	26	26	36
4	22	B14-C03	219	67	99	60	115	50	35	43	38	27	53	49	49	53	57	27	55	23	27	52	17	22
5	21	B14-H01	213	57	149	53	81	53	16	40	28	12	43	55	55	50	58	18	56	19	25	48	9	24
6	17	B14-S04	173	34	86	31	63	51	21	38	17	17	40	76	41	48	49	18	33	16	15	48	10	12
7	17	B06-H	173	54	46	53	41	38	32	36	33	37	30	32	32	30	17	28	22	25	25	19	31	16
8	16	B14-L06	165	56	44	39	40	30	38	25	37	32	35	25	27	26	23	29	20	24	31	18	19	13
9	16	B14-S01	165	67	83	48	60	53	7	16	22	10	81	36	29	34	51	16	47	15	34	48	5	30
10	16	B14-F07	158	29	80	35	70	49	10	35	15	7	35	42	42	75	46	9	45	9	7	44	8	11
11	15	B14-K01	147	47	77	29	49	41	22	31	34	14	33	30	42	38	36	21	32	16	26	36	7	23
12	14	B14-S16	147	54	56	38	48	34	34	38	42	24	36	32	36	32	26	33	32	23	28	31	8	15
13	14	B14-C01	147	73	40	52	39	26	58	49	85	43	26	23	33	15	16	57	17	45	31	22	15	28
14	14	B14-E12	144	58	44	44	34	35	58	90	44	37	12	34	45	32	7	45	11	33	16	10	10	21
15	14	B14-F02	137	43	70	36	51	39	23	51	29	19	26	39	60	42	23	21	24	16	18	20	9	14
16	14	B14-F01	135	40	69	41	50	30	21	45	27	20	26	41	67	39	25	27	15	21	25	7	17	
17	13	B07-H	135	28	36	42	29	25	21	31	17	27	22	28	31	28	16	14	15	14	15	16	18	7
18	13	B14-N1 0	130	31	78	35	48	39	8	18	13	6	27	42	39	34	40	9	30	6	13	31	6	13
19	13	B14-C09	127	34	75	29	57	31	10	27	22	7	32	33	39	40	9	36	6	15	32	5	14	
20	12	B14-N1 7	126	27	64	36	56	29	21	27	26	13	37	31	22	32	39	15	32	10	16	37	6	8
21	12	B14-N03	123	34	55	27	45	30	18	19	16	16	26	32	24	32	28	9	22	7	16	33	8	13
22	11	B14-J05	114	41	41	39	30	21	25	25	27	19	20	24	19	16	19	21	20	15	20	22	7	8

Derwent Innovation=Manual Code x  
Derwent Pharma = Therapy Area

どのような治療分野でどのようなマニュアルコード  
のエリアの特許が出ているのかを把握する

Reset		DWPI Manual Codes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
		# Records	50	38	37	36	33	31	28	17	17	15	13	12	12	11	11	11	11	10	10
		Show Values >= 1																			
		Cooccurrence # of Records																			
Actions	# Records		Neuroprotectant	Beta secretase inhibitor	Neotropic agent	Amyloid protein deposition inhibitor	Anticancer	CNS diagnostic agent	Anti-inflammatory	Beta amyloid antagonist	Gamma-secretase inhibitor	Diagnostic agent	p38 MAP kinase inhibitor	PAI inhibitor	Gene therapy	Antiparkinsonian	Antiviral	Neoplasm diagnostic agent	Glycogen synthase kinase-3 beta inhibitor	Antisense oligonucleotide inhibitor	Immunomodulator
1	29	B14-J01 A4	29	26	19	23	19	15	15	14	13	8	7	5	7	5	8	4	2	6	4
2	28	B14-N1 6	17	21	11	20	7	8	9	7	6	4	10	10	1	4	3	1	11	2	2
3	24	B14-J01	12	14	10	10	8	3	5	4	5	1	7	5	1	3	2		10	1	3
4	22	B14-C03	7	9	7	6	10	3	20	5	2	6	12	7	2	3	5	2	5	4	3
5	21	B14-H01	10	1	10	2	21	5	14	1	1	4	10	10	3	6	5	5	9	2	2
6	17	B14-S04	5	8	5	3	10	5	12	1	2	3	10	10	1	3	4	3	5	1	5
7	17	B06-H	7	20	3	13	8	3	3	5	7	2	6	2		3	2	1	2		2
8	16	B14-L06	16	6	13	11	8	7	7	7	4	6	2	6	5	3	2		4		2
9	16	B14-S01	17	4	8	4	11	6	16	2	2	4	12	8	2	4	6	3		4	2
10	16	B14-F07	5	8	2	1	10	2	12	1	2	6	10	10	2	1	2	3	2		2
11	15	B14-K01	3		2	1	7	1	10	1		4	9	10	1	1	5		4	2	1
12	14	B14-S16	7	2	3	3	6	10	1	1	1	6	6			2	5		3		2
13	14	B14-C01	5	2	7		7		4			7				4	1		2		3
14	14	B14-E12	3	2	6	7	1	2				1	3	8		1			4		2
15	14	B14-F02	1	1	6	2	4	3	6			1	3	10	10	1	2	1	1	5	1
16	14	B14-F01	2	1	4	2	4	2	6	1		4	10	9		2	1		4		4
17	13	B07-H	4	23	3	11	7	3	7	5	7	2				4	1	1	1	3	1
18	13	B14-N1 0	4	1	4	3	8	3	11	1		3	10	10		1	3	1	4		3
19	13	B14-C09	2	2	2	6	1	9	1			1	13	8		1	1	3	1		1
20	12	B14-N1 7	2		3		8	1	14	1						3	3	1	4		3
21	12	B14-N03	4		5	4	8	2	9			3	3	9	6		2	5		1	1

Derwent Innovation=Manual Code x  
Derwent Pharma = Action

どのような薬効の特許がどのようなマニュアルコード  
のエリアに出ているのかを把握する

# 機械学習機能を利用して類似特許を探す

# 類似特許の検索

## 類似特許の検索には2つのスタイルがあります。

類似特許の検索は、トレーニングフィールドを設定することにより、設定した範囲内で類似する特許を検索して、類似度スコアを付けて結果を表示する機能です。例えば、用途が似ている特許、優位性(効果)が似ている特許、発明内容が似ている特許など、観点を絞った類似特許の検索が可能です。

### DDA集合内に存在するレコードの特定の内容に類似する他のレコードを探す

ターゲットとなる特許等(複数可)と調査対象の集合を混ぜた集合を作成し、DDA内で、ターゲットに近い内容のレコードを探してもらう機能。  
侵害調査や監視対象の調査など、既存の特許などの類似特許を探す場合に適しています。

### DDA集合内に存在しない内容に類似する他のレコードを探す

文章(英文)を入力し、それに近い内容を集合から探してもらう機能。  
出願前調査など、未公開の文書の類似特許などを探す場合に適しています。

## 機械学習機能利用のための設定

類似特許の検索の機能を利用するには、機械学習をどのフィールドに基づいて行うのか？を設定する必要があります。

フィールド	項目数	グループ数	カバー率 (%)	データタイプ	メタタグ
(フィルター)					
タイトル	7,413		99%		
タイトル (ベスト)	8,402	1	100%		DDA 由来, トレーニングフィールド
タイトル (ベスト) (文章区切り) (フレーズ)	23,151	2	100%		DDA 由来
タイトル (ベスト) (文章区切り) (ワード)	8,458	5	100%		DDA 由来

- メタタグに、**トレーニングフィールド** として設定されている項目が、機械学習で利用されます。  
  
似ている特許レコードの発見や、分類を行うときの、指標となるフィールドを、トレーニングフィールドのタグを付与してください。  
  
フィールドで、右クリックして、「**メタタグを設定**」をクリックすると、設定できます。

## DDA集合内に存在するレコードの特定の内容に類似する他のレコードを探す

### ターゲットが1レコードの場合

1

タイトルウィンドウからダブルクリックしてターゲットのレコード画面を開きます。

2

フィールド名	値
公報番号 (Derwent Innovation リンク)	<a href="http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=US20160200950A1">http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=US20160200950A1</a>
DWPI アクセション番号	2013B02550
パーシク特許	CN102730715A / A
タイトル (ベスト)	Active mixture for st... m chloride powder, : hol which are mixed

3

類似レコードの最大数（上限）と最低一致率を入力します。最低一致率を高くすると、類似度の高いレコードがヒットします。

#### 類似レコードオプションを検索

検索する類似レコードの最大数:

10

続行

検索するレコードの最低一致率 (%):

13

☐ 類似レコードを検索するときに表示しない

4

### ターゲットが複数レコードの場合

1

タイトルウィンドウでターゲットのレコードを選択。

2

右クリックして「類似レコードを検索」をクリック。

類似レコードを検索

類似レコードを表示 (1 / 10) 28% 類似性

Field	値
公報番号 (Derwent Innovation リンク)	<a href="http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=US20160200950A1">http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=US20160200950A1</a>
DWPI アクセション番号	2016427248
パーシク特許	US20160200950A1 / A1 / 2016-07-14
タイトル (ベスト)	Adhesive composition useful in e.g. article, toys, musical instruments, window frames and sills, doors and flooring, comprises polymer blend comprising first propylene-based polymer, and second propylene-based polymer and propylene polymer
引用(Forward) 特許数-DPCI	1
DWPI ファミリー国数	131
DWPI ファミリーメンバー数	6
独立請求項 (数)	
優先権主張番号 (ロング)	
譲受人/出願人 (ベスト)	

類似のレコードがパーセンテージとともに、レコード画面の形式で表示されます。

## DDA集合内に存在しない内容に類似する他のレコードを探す

1. テキストに類似のレコードを検索

検索する類似レコードの最大数: 50

検索するレコードの最低一致率 (%): 30

テキストを入力

The acid-type carboxylated cellulose nanofibers are useful in spray compositions, rubber reinforcing materials, resin reinforcing materials, cosmetics, medical products, foods, beverages and paints. The acid-type carboxylated cellulose nanofiber has high viscosity and very short fiber length.

2. 条件、テキストなどを入力

3. OK

検索結果 (195 タイトル、1 選択)

- Acid-type carboxylated cellulose nan...
- Acid-type carboxymethylated cellul...
- Additive used e.g. as cosmetic additi...
- Additive used in foam-containing c...
- Additive used in shape-retaining ag...
- Anion-modified cellulose nanofiber ...
- Antifog additive used for anti-fog fil...
- Aqueous coating material for formin...
- Aqueous composition used as additi...
- Aqueous dispersion liquid for formi...
- Aqueous resin composition used in ...
- Artificial casing used for storing pro...
- Baked foodstuff dough for baking b...
- Base paper for processing papers e....
- Base paper used in release paper for ...
- Binder for ceramics used for formin...
- Building material used as mortar an...
- Carboxymethylated cellulose nanofi...
- Carboxymethylated cellulose nanofi...
- Carboxymethylated cellulose nanofi...
- Carboxymethylated cellulose nanofi...

類似レコードを表示

類似レコードの表示 (1 / 3) 60% 類似性

すべて表示

Field	値
公報番号 (Derwent Innovation リンク)	<a href="http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=WO2018116661A1">http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=WO2018116661A1</a>
DWPI アクセッション番号	201850977D
パテント特許	WO2018116661A1 / A1 / 2018-06-28
タイトル (ベスト)	Acid-type carboxylated cellulose nanofibers useful e.g. in spray compositions, rubber reinforcing materials, resin reinforcing materials and paints, comprise carboxylated cellulose nanofiber having carboxyl group

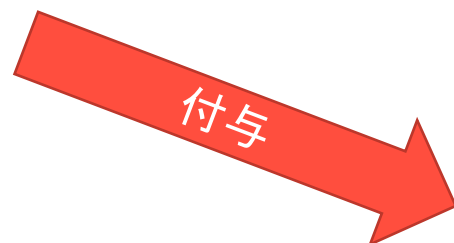
類似のレコードがパーセンテージとともに、レコード画面の形式で表示されます。

# 機械学習機能を利用して自動分類を行う

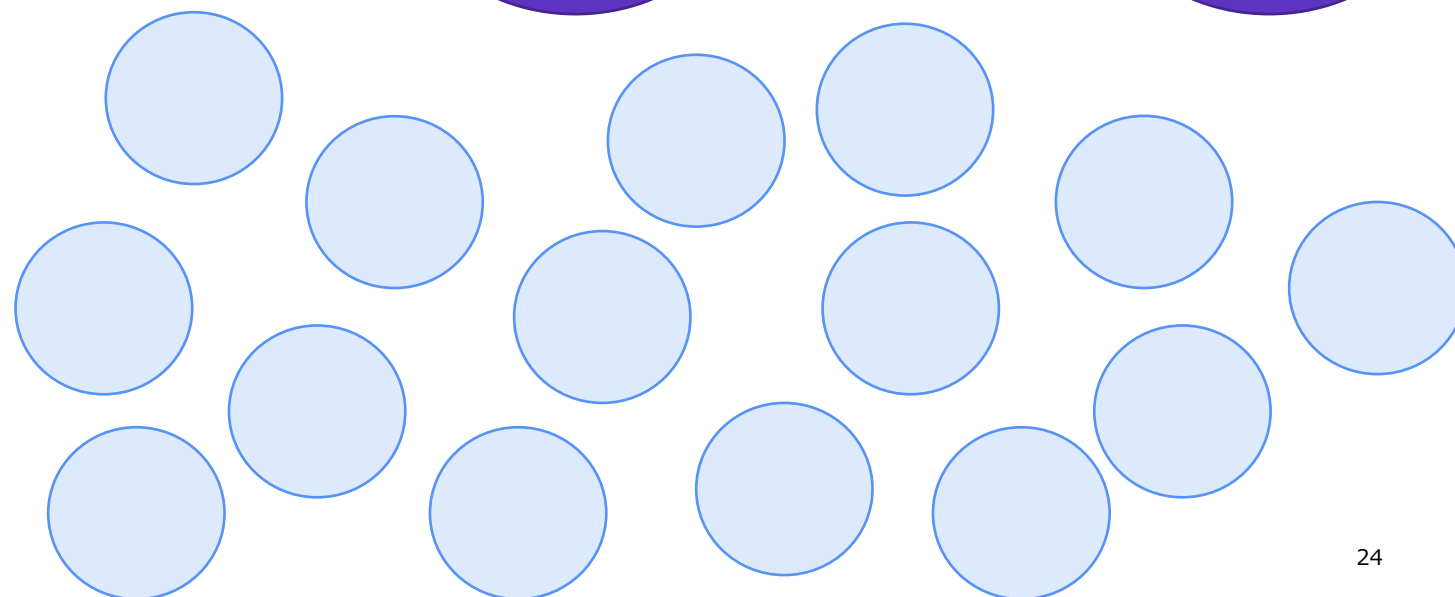
# Derwent Data Analyzer:自動分類の仕組み



機械学習により、カテゴリごとの  
ワード/分類の傾向などを  
トレーニングフィールドの設定  
に応じて把握



学習結果を活かして  
まだ未分類のレコードに  
カテゴリを付与する



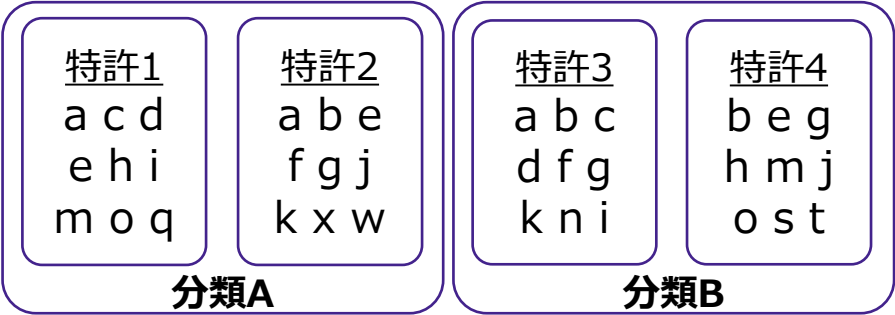


# Derwent Data Analyzer:自動分類の仕組み

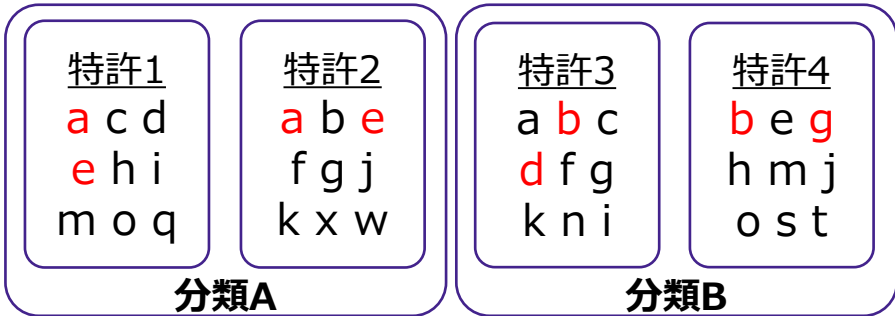
## 1. 機械学習プロセス

ユーザーが用意した教師データ

分類分け（分類付与）した特許のテキスト（a,c,b :  
公報、抄録などのワード）

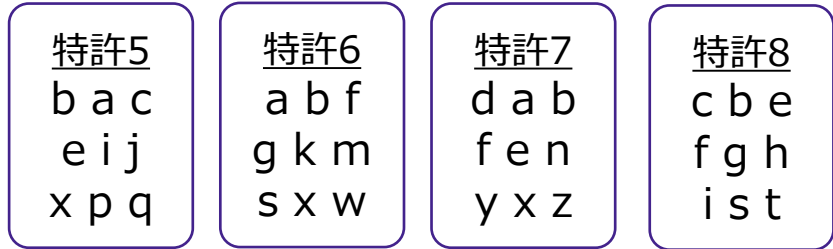


Derwent Data Analyzerが  
機械学習により分類毎のワードの傾向を把握

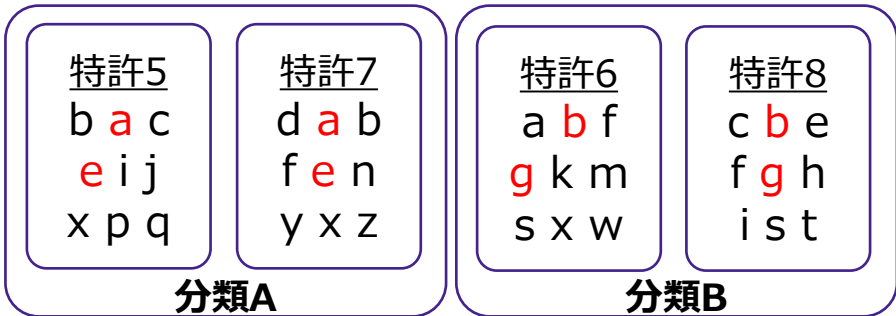


## 2. 分類分け（分類付与）プロセス

未分類の特許



学習により把握した傾向と同じような傾向を持つ特許を分類分け



学習結果を活用

# DDAで教師データを作成して、 自動分類を行う

# DDAで教師データを作成して、自動分類を行う



DDA上で、レコードを分類分けするような作業をする場合、DDA上で独自にフィールドを作成し、分類項目を作成、各レコードを分類に振り分けていくことが可能です。

更に、一部ユーザーが手動で振り分けた分類の傾向を機械学習により学習し、残りのレコードの分類を自動的に行う機能も付加されています。

## 1 分類フィールドを作成



## 2 トレーニングフィールドの設定



## 3 レコード画面で手動でカテゴリを付与(教師データの作成) AとBの方法



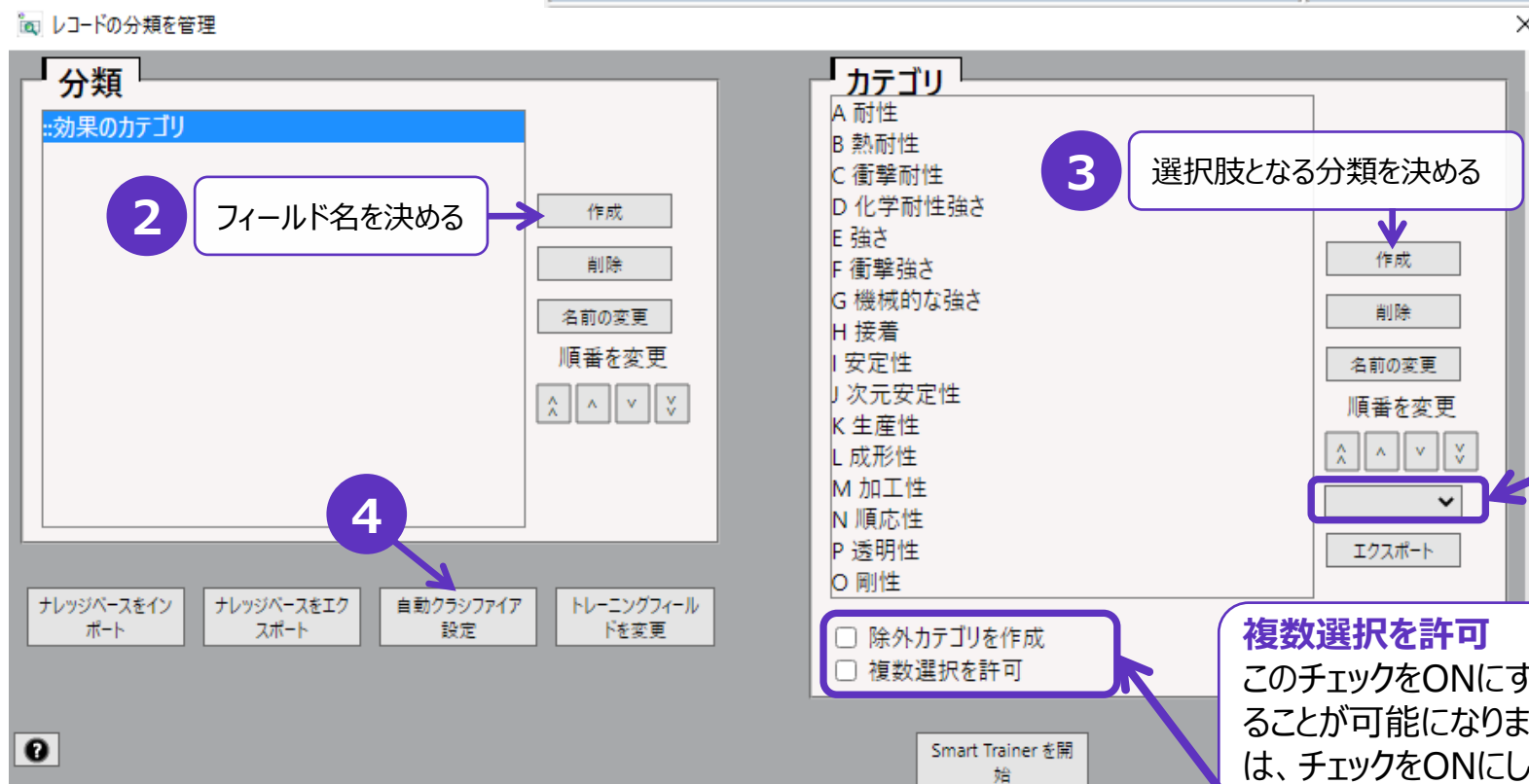
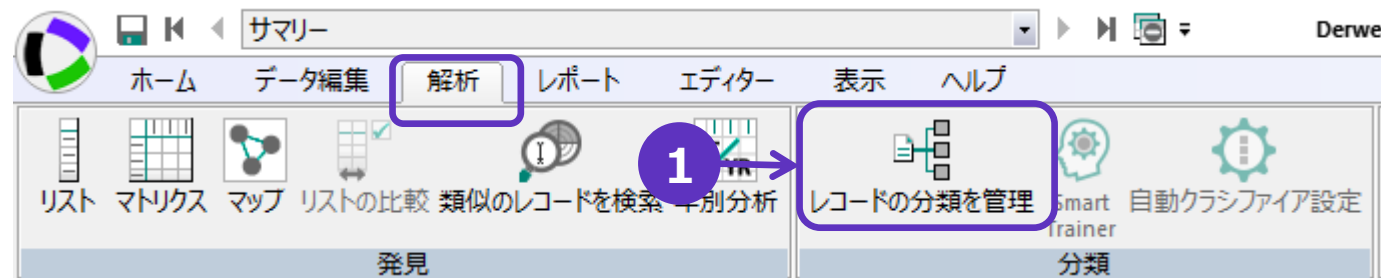
(DDAが自動で機械学習していきます…)



## 4 残りのレコードに対して自動で分類を付与する

# DDAで教師データを作成して、自動分類を行う

## 1 分類フィールドを作成



分類項目がたくさんある場合は、インポート機能を利用します。  
テキストファイルに、分類を改行区切りで入力したファイルを作成しておきます。  
(※日本語を含む場合は、unicodeで保存)

### 複数選択を許可

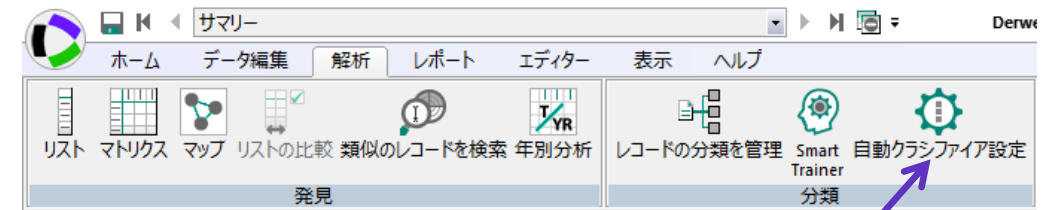
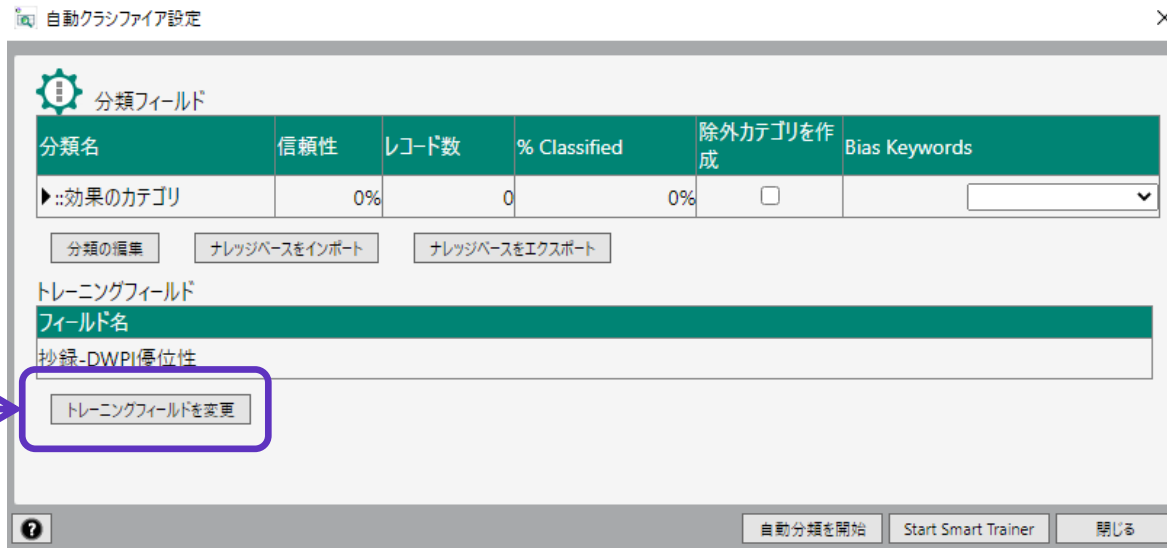
このチェックをONにすると、1つのレコードに対し、複数の分類を付与することが可能になります。複数の分類に当てはまる可能性がある場合は、チェックをONにしておきます。

### 除外カテゴリを作成

「除外カテゴリ」は教師データの中に学習させたくないものがある場合に、それらを無視するものとして印を付けるカテゴリです。Disregardと表示されます。

# DDAで教師データを作成して、自動分類を行う

## 2 トレーニングフィールドの設定



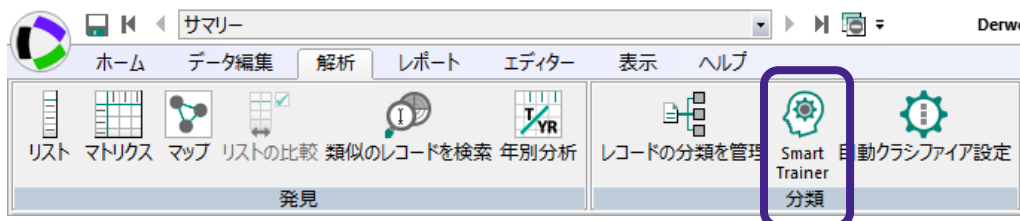
ココからでも同じ画面を開けます。



# DDAで教師データを作成して、自動分類を行う

## 3 レコード画面で手動でカテゴリを付与（教師データの作成）

### A. Smart Trainerを利用する



ユーザーの手作業をDDAが機械学習していきます・・・

ユーザーによる分類の付与が始まると、その内容にしたがい、自動的に機械学習が始まります。付与を行えば行うほど、信頼度が向上します。

7 自動付与へ

6 次へ

5 内容を読みながらカテゴリ付与

トレーニングフィールドに指定されている部分のデータが表示されるので、そこを確認しながら、手動でカテゴリ付与を行っています。

効果 116/207

- ☐ A 耐性
- ☐ I 安定性
- ☐ K 生産性
- ☐ E 強さ
- ☐ B 熱耐性
- ☐ H 接着
- ☐ J 次元安定性
- ☐ L 成形性
- ☒ G 機械的な強さ
- ☐ C 衝撃耐性
- ☐ M 加工
- ☐ N 順応
- ☐ P 透明性
- ☐ D 化学耐性
- ☐ F 衝撃強さ
- ☐ O 剛性
- ☐ Disregard

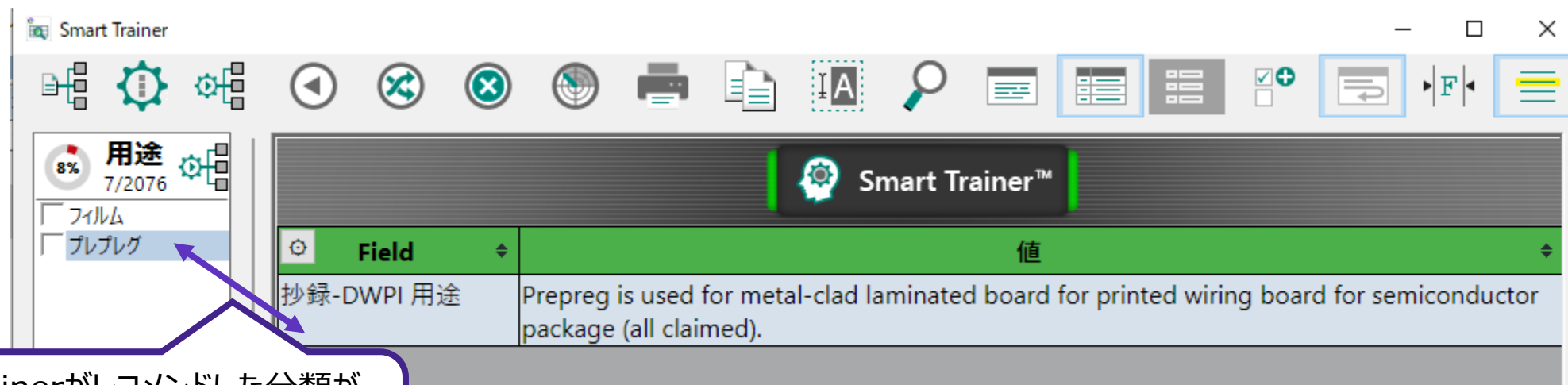
Field	値
抄録-DWPI優位性	Separation of protective layer from 1st polymeric layer is easy without separating 1st polymeric layer from carrier substrate. High glass transition temperature of baked 1st polymeric layer can help maintain mechanical integrity of resulting devices.

このレコードに関するノート

☐ 新しいデータセットから除く

## 分類のレコメンド：より正確に自動分類できるようにSmart trainerが適当な分類をリアルタイムでレコメンド（推薦）

- ◆ 教師データの分類を手動で作成する際、Smart trainerが適当な分類をリアルタイムでレコメンド（推薦）してくれます。これにより、Smart trainerが正しく学習できているかどうか確認することができます。
- ◆ 適切な分類がレコメンドされる場合、特徴を捉えて正しく学習していると判断できます。適切な分類がレコメンドされない場合、誤って学習していると判断できます。その場合はそれまでの手動による分類作業の見直しや分類の追加など対応が必要になります。



Smart trainerがレコメンドした分類が  
ハイライトされます。  
正しく学習しているかどうか目安になります。

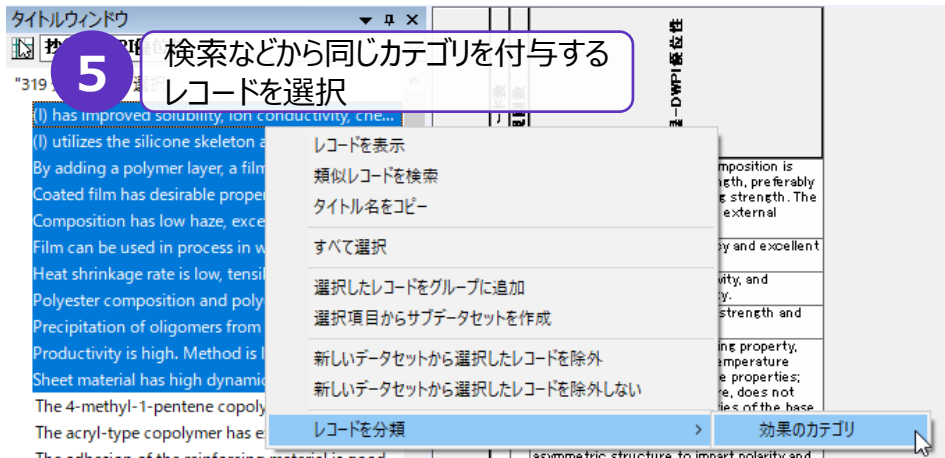
# DDAで教師データを作成して、自動分類を行う

## 3 レコード画面で手動でカテゴリを付与（教師データの作成）

### B. レコード画面を使って行う

ユーザーの手作業をDDAが機械学習していきます....

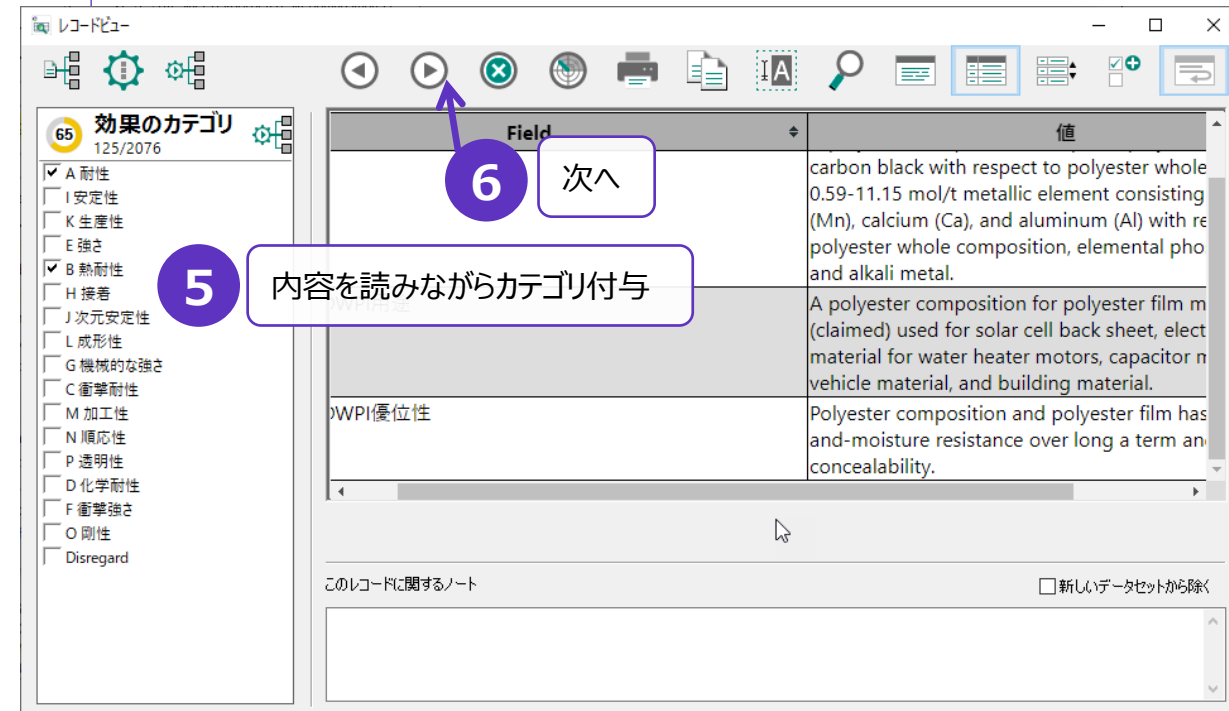
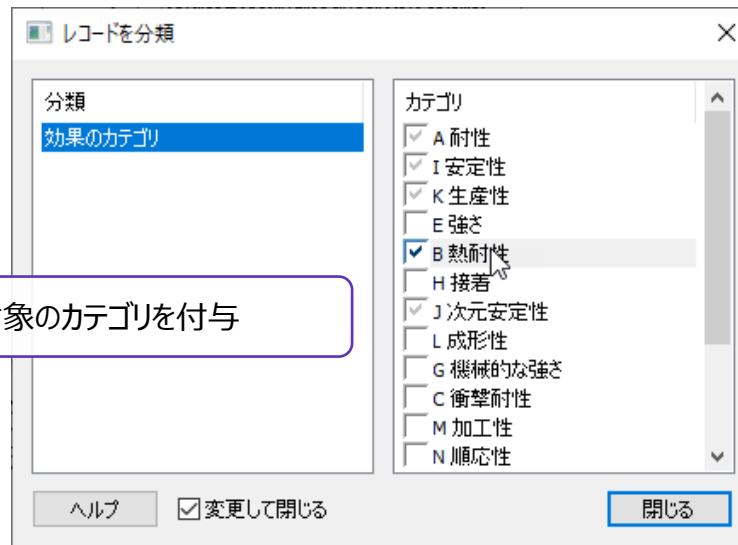
ユーザーによる分類の付与が始まると、その内容にしたいがい、自動的に機械学習が始まります。付与を行えば行うほど、信頼度が向上します。



【複数に一気に付与】

【レコード画面で1レコードずつ付与】

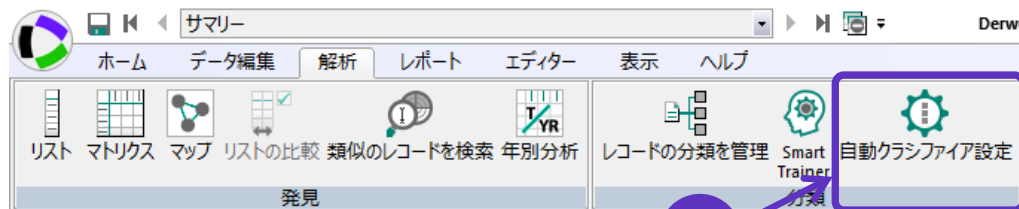
6 対象のカテゴリを付与





# DDAで教師データを作成して、自動分類を行う

## 4 残りのレコードに対して自動で分類を付与する



### 信頼度を確認

この数値は、機械学習の結果、どの程度、カテゴリ間の傾向が明らかになったか？を示します。  
※この信頼度を上げるには、トレーニングフィールドを変更して学習しなおすか、または教師データを見直します。

### 精度目標

精度低めでよりたくさんのレコードをカテゴリに組み込むか、精度を高くし、Uncategorizedに入るものが増えてもよいかの選択です。  
精度を高めにするをお勧めします。

自動クラシファイア設定

分類フィールド

分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::効果のカテゴリ	17	66%	6%	<input checked="" type="checkbox"/>

分類の収集 ナレッジベースをインポート ナレッジベースをエクスポート

自動クラシファイア設定

カテゴリ	レコード数	信頼性
Disregard	2	0%
A 耐性	54	80%
I 安定性	42	73%
K 生産性	37	67%
E 強さ	36	54%
B 熱耐性	33	55%
H 接着	20	58%

自動分類を開始 Smart Trainerを開始 閉じる

自動的にレコードを分類

自動分類処理のための分類フィールドを選択

信頼性	分類済み	分類名
<input checked="" type="checkbox"/> 65%	6.0%	効果のカテゴリ

すべて選択 選択なし アルファベット順にソート

自動クラシファイアの進捗

6 %

分類を開始 ヘルプ 閉じる

精度目標

高いより少ないレコードを分類することで、分類の誤りが少ない

低いより多いレコードを分類することで、分類の誤りが多い

## 7 必要に応じて自動付与データをダウンロードまたは再分類

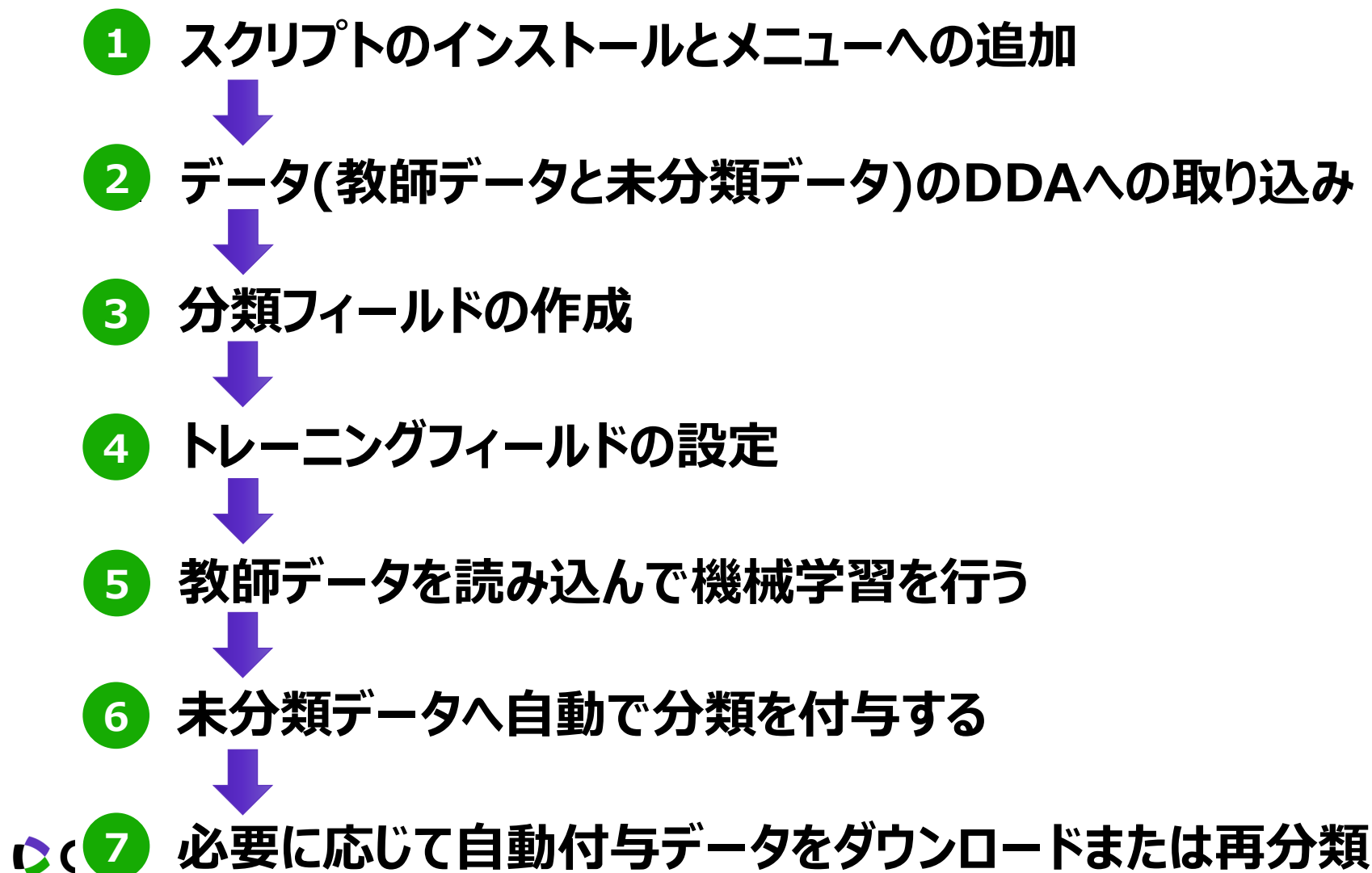
ダウンロードの方法については、「様々な処理」-「Excelへのデータ出力」を参照してください。  
再分類の方法については、「自動分類結果を用いた再分類」を参照してください。

既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

既に、自社でカテゴリわけした分類とレコードがある場合は、それをDDAに取り込んで機械学習させることも可能です。

手動で行うのではなく、自動的に取り込んで学習させるスクリプトが用意されていますので、そのスクリプトを使って取り込みを行う方法をご紹介します。

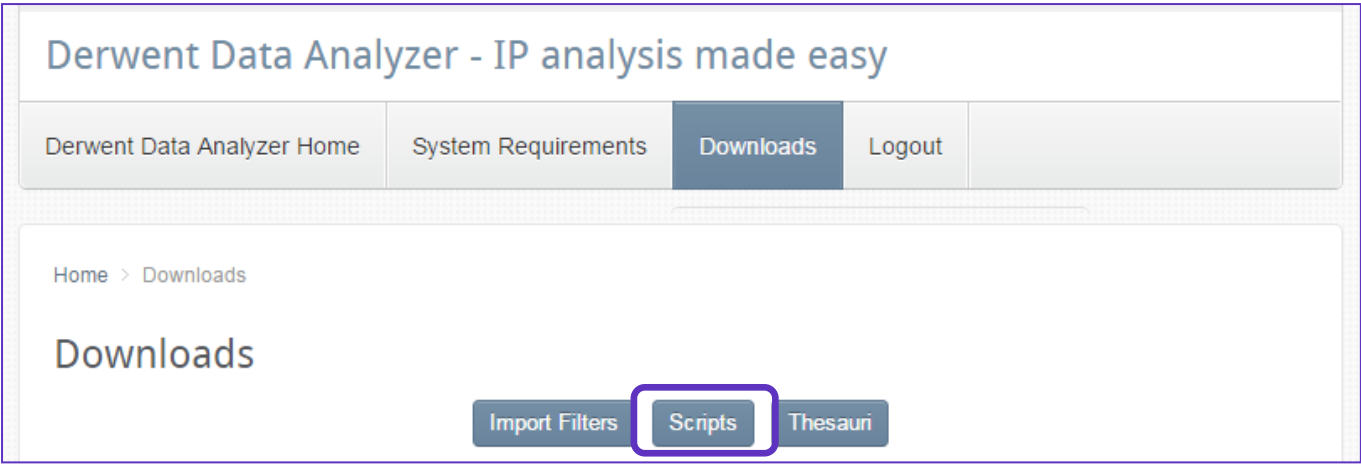


# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 1 スクリプトのインストールとメニューへの追加



ダウンロードページが表示されます。

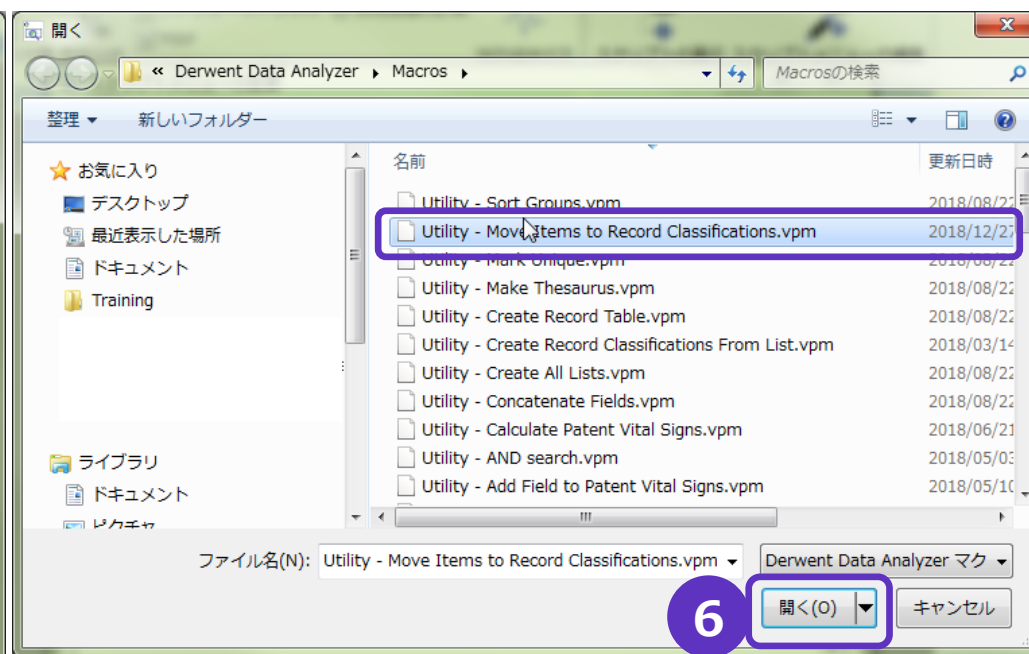
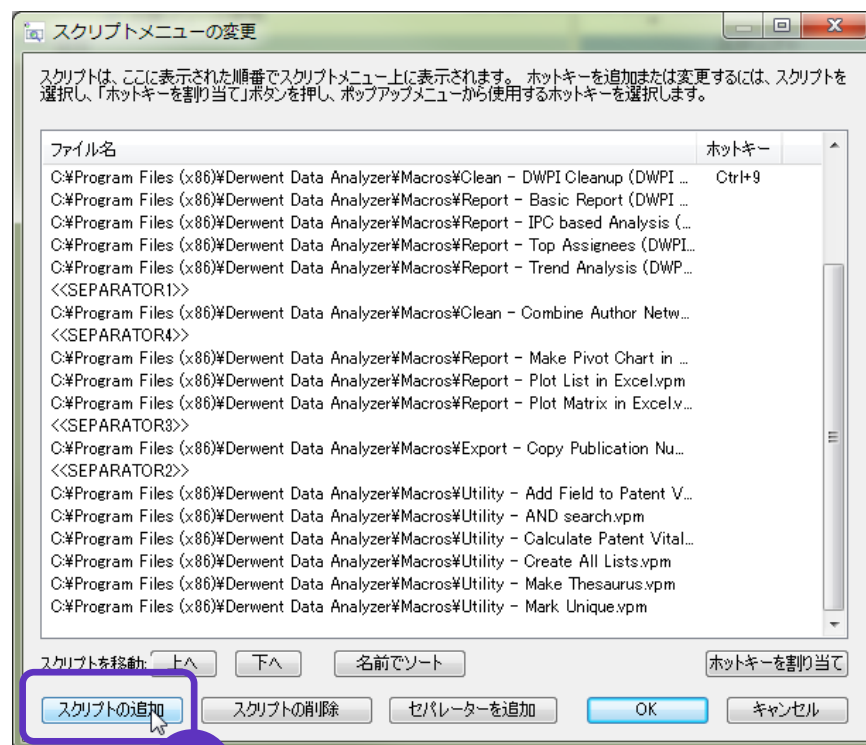
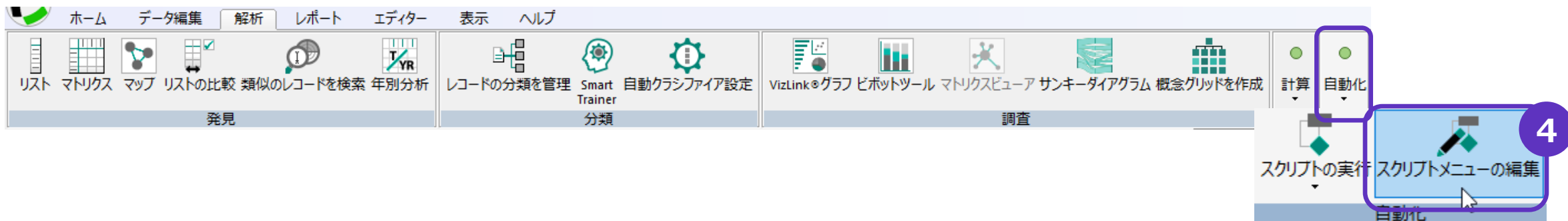


ダウンロードページが表示されます。

Name	Filename	Last Update
Utility - Move Items to Record Classifications	Utility - Move Items to Record Classifications.zip	02/07/2019
	<div>Install</div>	
Requirements:	New Record Classification field is already created before running the script.	

3 Utility – Move Items to Record Classifications.zip  
を選択してインストール

# スクリプトをメニューに追加します



先ほどダウンロードした  
Utility - MoveItems to Record Classifications.vpm  
を指定して「開く」

スクリプトをメニューに追加、位置を決めます。  
上へ、下へのボタンを使い、わかりやすいところに配置します。また、セパレータを入れておくと分かりやすいです。

スクリプトメニューの変更

スクリプトは、ここに表示された順番でスクリプトメニュー上に表示されます。ホットキーを追加または変更するには、スクリプトを選択し、「ホットキーを割り当て」ボタンを押し、ポップアップメニューから使用するホットキーを選択します。

ファイル名	ホットキー
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Report - IPC based Analysis (...)	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Report - Top Assignees (DWPI...	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Report - Trend Analysis (DWP...	
<<SEPARATOR1>>	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Clean - Combine Author Netw...	
<<SEPARATOR4>>	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Report - Make Pivot Chart in ...	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Report - Plot List in Excel.vpm	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Report - Plot Matrix in Excel.v...	
<<SEPARATOR3>>	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Export - Copy Publication Nu...	
<<SEPARATOR2>>	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - Add Field to Patent V...	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - AND search.vpm	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - Calculate Patent Vital...	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - Create All Lists.vpm	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - Make Thesaurus.vpm	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - Mark Unique.vpm	
<<SEPARATOR5>>	
C:\Program Files (x86)\Derwent Data Analyzer\Macros\Utility - Move Items to Record...	

スクリプトを移動: 上へ 下へ 名前ですりて

ホットキーを割り当て

スクリプトの追加 スクリプトの削除 セパレーターを追加 OK キャンセル

Visual Basic スクリプトの実行

レポート - 企業比較	Ctrl+1
レポート - テクノロジーレポート	Ctrl+2
レポート - 企業レポート	Ctrl+3
レポート - DDA ピボットグラフ	
レポート - 特許維持状況	
クリーン - DWPI クリーンアップ (DWPI のみ)	Ctrl+9
レポート - 基本レポート (DWPI のみ)	
レポート - IPC ベースの分析 (DWPI のみ)	
レポート - 上位の譲受人/出願人 (DWPI のみ)	
レポート - 傾向分析 (DWPI のみ)	
クリーン - 著者ネットワークの結合	
レポート - ピボットグラフを Excel で作成	
レポート - リストを Excel でプロット	
レポート - マトリクスを Excel でプロット	
エクスポート - 公報番号をクリップボードにコピー	
ユーティリティ - 特許維持状況にフィールドを追加	
ユーティリティ - AND 検索	
ユーティリティ - 特許維持状況の計算	
ユーティリティ - すべてのリストの作成	
ユーティリティ - シソーラスの作成	
ユーティリティ - 一意のマーク	
Utility - Move Items to Record Classifications	



# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 2 データ(教師データと未分類データ)のDDAへの取り込み

分類済みのデータと未分類のデータを一緒にしたExcelファイルの作り方についてポイントをご紹介します。

行にデータが並んでいる状態で作成してください。

教師データとなる情報と未分類のデータを1つのファイルにしておくインポートが一回で済みます。

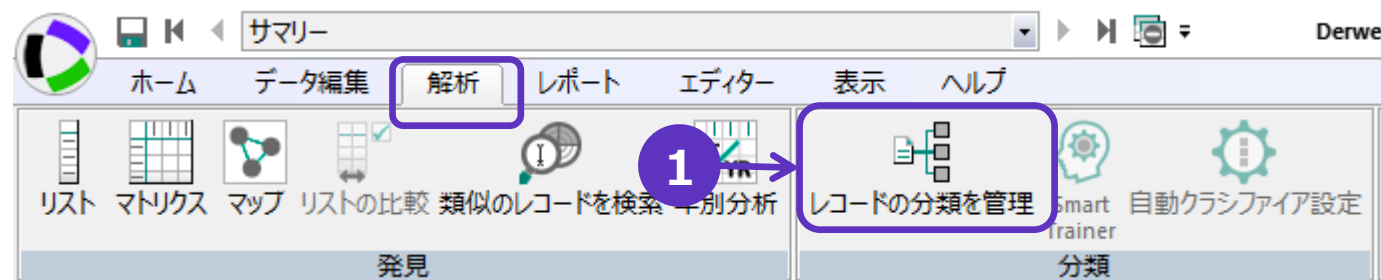
(※教師データと未分類のデータでファイルを分けてある場合、別々に取り込んだ後、マージする方法でも行うことができます(データセットの融合)。)

分類済みの教師データについては、  
既存の分類を入力した状態にします。

	A	B	C	D	E	F	G
1	公報番号	抄録-DWPI新規性	抄録-DWPI新規性の分類	抄録-DWPI優位性	抄録-DWPI優位性の分類	抄録-DWPI用途	抄録-DWPI用途の分類
2	JP2013139523A	A polybutylene terephthalate film; H樹脂		Polybutylene terephthalate film; H樹脂	安定性; K生産性; J次元安定性	A polybutylene terephthalate film; F建材; K包装	
3	WO2013105340A1	A carbon fiber reinforced plastic; G繊維; H樹脂		The carbon fiber has large diameter; A耐性; E強度; C衝撃耐性		Carbon fiber reinforced plastic; A自動車; Eモバイル; H飛行機	
4	WO2013105300A1	Aromatic polyamide film; Aフィルム; Jポリアミド		The short circuit of the polyamide film; A耐性; B熱耐性		Aromatic-polyamide film; A自動車	
5	KR2013081453A	A white opaque polyester film; Aフィルム; Bポリエステル; F組成物; H樹脂		The white opaque polyester film; E強度		White opaque polyester film; Dフィルム	
6	KR2013081124A	A heat-shrinkable polyester film; Aフィルム; Bポリエステル; H樹脂		The heat-shrinkable polyester film; A耐性		Heat-shrinkable polyester film; Dフィルム	
7	KR2013080028A	Polyimide film (e), is a film; Aフィルム; Iポリイミド		The polyimide film: has ultrahigh strength; A安定性; K生産性; J次元安定性		The polyimide film is used for; Dフィルム	
8	KR2013076391A	Manufacture of meta-aramid film; Aフィルム		The meta-aramid-based film; E強度		Manufacture of meta-aramid film; C製造; Dフィルム	
9	KR2013076390A	Meta-aramid film containing; Aフィルム		The film exhibits excellent mechanical strength; E強度; G機械的な強度		The meta-aramid film is used for; Dフィルム	
10	KR2013076044A	An adhesive protective film; Aフィルム; F組成物; H樹脂		The adhesive protection film; A耐性; H接着		Adhesive protection film; Dフィルム	
11	KR2013074955A	Copper-clad laminate; Aフィルム; Cエポキシ樹脂; F組成物; H樹脂		The copper-clad laminate; N順応性		Used as a copper-clad laminate; Dフィルム	
12	FR2985263A1	Recycling a multilayered film; Aフィルム		The process: produces recycled film; A耐性; I安定性; K生産性; D化		The process is useful for; Dフィルム; F建材	
13	FR2985261A1	The process comprises; Aフィルム; Bポリエステル		The process enables efficient production; A耐性; I安定性; K生産性; J次元安定性		The process is useful for; Dフィルム	
14	WO2013099741A1	A carbon fiber based material; Dカーボン; G繊維		The carbon fiber based material; E強度		Carbon fiber based material; A自動車; Bプレプレグ; H飛行機	
15	WO2013099608A1	A biaxially-oriented polyester film; Aフィルム; Bポリエステル		The biaxially-oriented polyester film; A耐性; I安定性; K生産性; L成形		Biaxially-oriented polyester film; A自動車; Dフィルム; Eモバイル	
16	WO2013099234A1	A polyphenylene sulfide film; F組成物; H樹脂		The polyphenylene sulfide film; A耐性; I安定性; K生産性; L成形		Polyphenylene sulfide film; A自動車	
17	WO2013099707A1	A sizing agent coated carbon fiber; Dカーボン; G繊維		The sizing agent coated carbon fiber; A耐性; H接着; I安定性; K生産性		Sizing agent coated carbon fiber; A自動車; Bプレプレグ; G船舶	
18	WO2013099862A1	An epoxy resin composition; Cエポキシ樹脂; F組成物; H樹脂		The epoxy resin composition; A耐性; E強度; B熱耐性		Epoxy resin composition; A自動車; Bプレプレグ; F建材	
19	WO2013099111A1	A thermoplastic resin composition; Bポリエステル; F組成物; H樹脂		The molded article obtained; A耐性; I安定性; B熱耐性		Thermoplastic resin composition is used for laser-marking; A自動車	
20	WO2013099607A1	A polyolefin microporous film; Aフィルム; H樹脂		The polyolefin microporous film; A耐性; E強度		Polyolefin microporous film; Dフィルム	
21	WO2013099539A1	A polyolefin microporous film; Aフィルム		The polyolefin microporous film; K生産性		Polyolefin microporous film; Dフィルム; Eモバイル; Jコン	

# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 3 分類フィールドの作成



レコードの分類を管理

分類項目がたくさんある場合は、インポート機能を利用します。  
テキストファイルに、分類を改行区切りで入力したファイルを作成しておきます。  
(※日本語を含む場合は、unicodeで保存)

### 複数選択を許可

このチェックをONにすると、1つのレコードに対し、複数の分類を付与することが可能になります。複数の分類に当てはまる可能性がある場合は、チェックをONにしておきます。

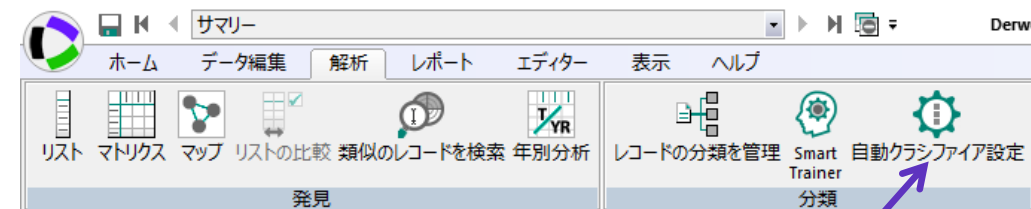
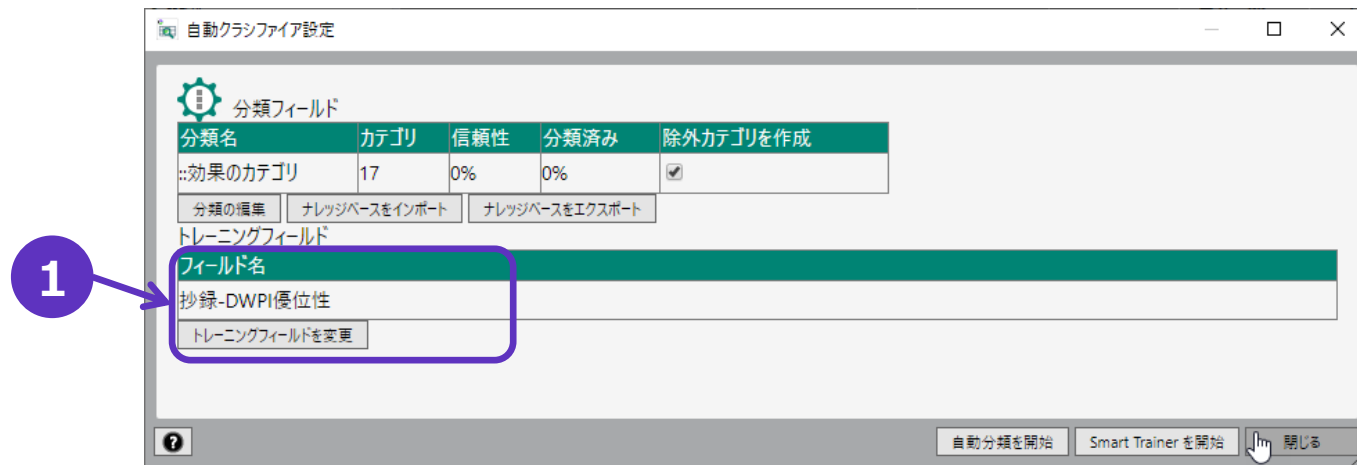
### 除外カテゴリを作成

「除外カテゴリ」は教師データの中に学習させたくないものがある場合に、それらを無視するものとして印を付けるカテゴリです。Disregardと表示されます。

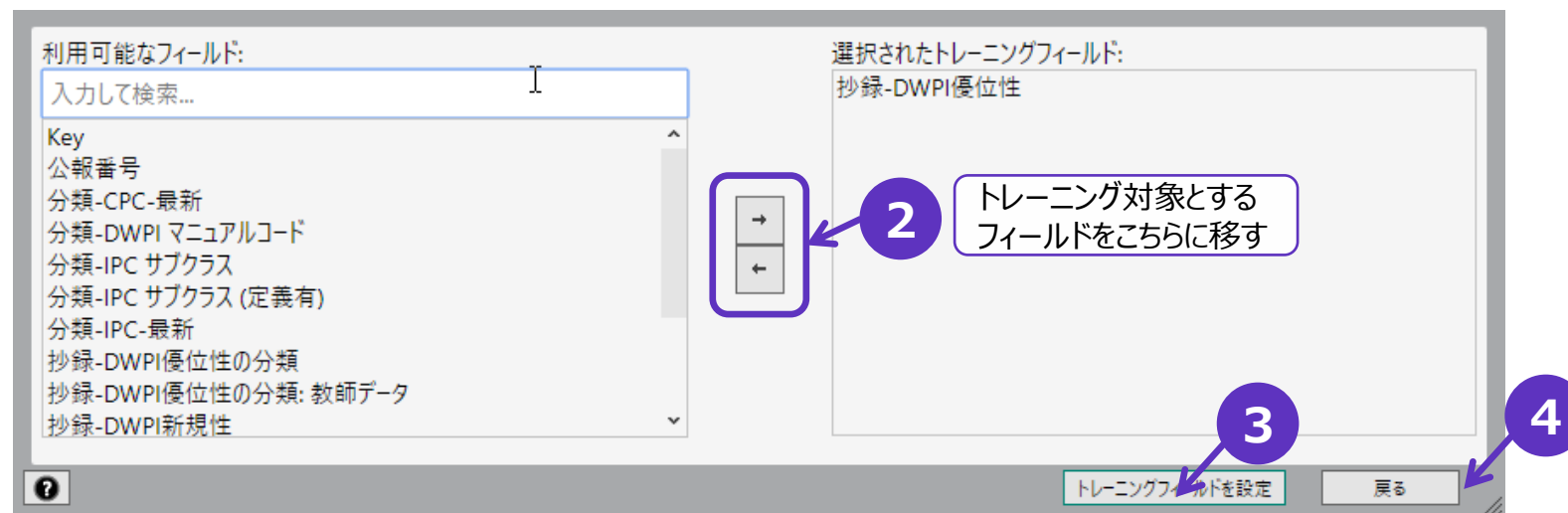


既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 4 トレーニングフィールドの設定



ココからでも同じ画面を開けます。



# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 5 教師データを読み込んで機械学習を行う

# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

トレーニングフィールドに基づき、機械学習が行われ、教師データになっているレコードが各分類のところにチェックがついた状態になります。

	レコード数		::効果のカテゴリ																	
	出現回数			A 耐性	I 安定性	K 生産性	E 強さ	B 熱耐性	H 接着	J 次元安定性	L 成形性	G 機械的な強さ	C 衝撃耐性	M 加工性	N 順応性	P 透明性	D 化学耐性	F 衝撃強さ	O 剛性	Disregard
1	1	1	14b7a4t3v3r4ck4w59516ak4m26ru2n5o	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>													
2	1	1	14hi5q716z1h5d3q5f6t41633n226r533o6k	<input checked="" type="checkbox"/>																
3	1	1	1aty4n4e204d2360245g5s6py25z3a2b41g	<input checked="" type="checkbox"/>																
4	1	1	1f48562i21211gb674q521vl15o26e1h3r1064	<input checked="" type="checkbox"/>											<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
5	1	1	1v2k6k3yo37415n31a48124u281e716m5b6i20	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>														
6	1	1	1w20485j4w5e2o1n6h363w36l1f5x333i34f5	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>						<input checked="" type="checkbox"/>							
7	1	1	235b2f711k3i3m593wk2e2w591al5t2i331n35	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>												
8	1	1	25513flh3kn32541u5o445h6wlo1u3g5puk2y	<input checked="" type="checkbox"/>																
9	1	1	2a3o673wl1n172kl15h1320534f5b6s2jc6p324s	<input checked="" type="checkbox"/>																
10	1	1	2fp4a5e7q3r4m6hb302f4i5f5g1y3h322a6w	<input checked="" type="checkbox"/>																
11	1	1	2h131m2r1i5f4z256i6j2i2o6s1f5n3v2rx713s	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>														
12	1	1	2r4n4au1mla52yl1i3f2is4h6m705i4p506x6u	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>													

レコード  
赤→黄  
緑になった

※精度  
見直して

自動クラシファイア設定

分類フィールド

分類名	カテゴリ	信頼性	分類済み	除外カテゴリ
::効果のカテゴリ	17	66%	6%	<input checked="" type="checkbox"/>

レコード画面を開くと、機械学習結果の信頼性も確認することができます。  
赤→黄→緑 の順に信頼性が色分けされています。  
緑になっていれば、かなりの信頼ある学習結果が得られていることを示します。

※精度を上げるためには、トレーニングフィールド、または教師データを見直してみます。

自動クラシフィア設定

分類フィールド

分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::効果のカテゴリ	17	66%	6%	<input checked="" type="checkbox"/>

分類の標準    ナレッジベースをインポート    ナレッジベースをエクスポート

::効果のカテゴリ カテゴリ

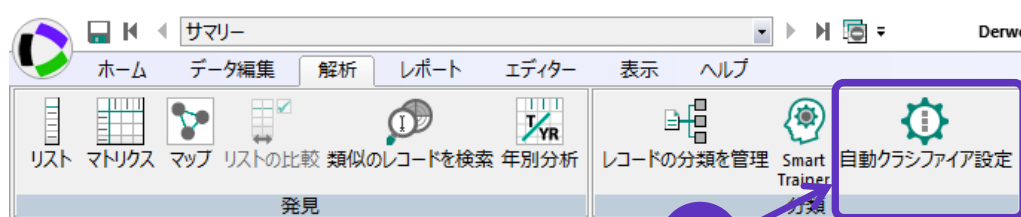
カテゴリ	レコード数	信頼性
Disregard	2	0%
A 耐性	54	80%
I 安定性	42	73%
K 生産性	37	67%
E 強さ	36	54%
B 熱耐性	33	55%
H 接着	20	58%

自動分類を開始    Smart Trainer を開始    閉じる

# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

## 6 未分類データへ自動で分類を付与する

機械学習した内容で、自動で未分類のレコードを分類にカテゴライズします。  
多くの手間がかかっていた分類付与を機械学習によって大幅に短縮することが可能です。



### 信頼度を確認

この数値は、機械学習の結果、どの程度、カテゴリ間の傾向が明らかになったか？を示します。  
※この信頼度を上げるには、トレーニングフィールドを変更して学習しなおすか、または教師データを見直します。

### 精度目標

精度低めでよりたくさんのレコードをカテゴリに組み込むか、精度を高くし、Uncategorizedに入るものが増えてもよいかの選択です。  
精度高めにすることをお勧めします。

The screenshot shows the '自動クラシファイア設定' dialog box. It has two main sections: '分類フィールド' (Classification Fields) and '自動的にレコードを分類' (Automatically Classify Records).

**分類フィールド (Classification Fields):**

分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::効果のカテゴリ	17	66%	6%	<input checked="" type="checkbox"/>

Buttons: 分類の振集, ナレッジベースをインポート, ナレッジベースをエクスポート

**自動的にレコードを分類 (Automatically Classify Records):**

自動分類処理のための分類フィールドを選択

信頼性	分類済み	分類名
<input checked="" type="checkbox"/> 65%	6.0%	効果のカテゴリ

Buttons: すべて選択, 選択なし, アルファベット順にソート

精度目標: 高いより少ないレコードを分類することで、分類の誤りが少ない (indicated by a red arrow pointing to the slider).

自動クラシファイアの進捗: 6% (indicated by a red arrow pointing to the progress bar).

Buttons: 分類を開始, ヘルプ, 閉じる

Bottom buttons: 自動分類を開始, Smart Trainerを開始, 閉じる

# 既存の分類データを利用（教師データの作成はDDA外で行う）して、自動分類を行う

自動分類が終了し、すべてのレコードに対し、分類が付与された状態になります。  
どこにも属さないと思われるものには、[Unclassified]が付与されています。

レコード数	出現回数	改善点自動分類	A 耐性	I 安定性	K 生産性	E 強さ	B 熱耐性	H 接着	J 次元安定性	L 成形性	G 機械的な強さ	C 衝撃耐性	M 加工性	N 順応性	P 透明性	D 化学耐性	F 衝撃強さ	O 剛性	Unclassified
37	1	14516n4h2d5w294u4c2z1d415u5q401k6c1m1j3l																	
38	1	14544nk4x1l3m124h6p3s5n1512c1p39k6e2c																	
39	1	145b1w6m656sy5w2m5c2g5h2fp1d133b565mi																	
40	1	146k712b1f524l6z1k2yl1q2q3k466n311v																	
41	1	146m534f2f2g3z5247i3t4y2g4m115r5o3p3i2n																	
42	1	14bfa4t3v3r4ck4w59516ak4m26ru2n5o																	
43	1	14hi5q716z1h5d3q5f6t41633n226r533o6k																	
44	1	14v4o5u3c6l113i5i5n6i4r6d2w1i2y46182j																	

レコードビュー

27 改善点自動分類

☐ A 耐性

☐ I 安定性

☐ K 生産性

☐ E 強さ

☒ B 熱耐性

☐ H 接着

☐ J 次元安定性

☐ L 成形性

☐ G 機械的な強さ

☐ C 衝撃耐性

☐ M 加工性

☐ N 順応性

☐ P 透明性

☐ D 化学耐性

☐ F 衝撃強さ

☐ O 剛性

☐ Unclassified

フィールド名	値
公報番号	JP2018009197A
抄録-DWPI新規性	A thermoplastic resin composition comprises 50-85 %mass polycarbonate resin (A), 0.1-20 %mass polyester resin (B), 0.1-20 %mass rubber-reinforced graft resin (C) obtained by polymerizing a vinylic monomer containing an aromatic vinyl compound in the presence of a diene-based rubber, a hydrogenated diene-based rubber, acrylic rubber, or ethylene-α -olefin-type rubber, and 0.5-12 %mass compound (D) containing a polyorganosiloxane polymer portion and a vinyl-type (co)polymer portion, with respect to total content of components contained in the resin composition.
抄録-DWPI新規性の分類	
抄録-DWPI優位性	The thermoplastic resin composition provides molded article with excellent damage resistance, heat resistance, cold shock resistance, coloring property, and external appearance, even in an airtight environment and an outdoor environment.
抄録-DWPI用途	Thermoplastic resin composition is used for forming a molded article

レコード画面を開くと、文章を読みながら、付与された分類を確認することができます。

## 7 必要に応じて自動付与データをダウンロードまたは再分類

ダウンロードの方法については、「様々な処理」-「Excelへのデータ出力」を参照してください。  
再分類の方法については、「自動分類結果を用いた再分類」を参照してください。

# 自動分類結果を用いた再分類

## 自動分類結果による再分類 -概要-

リセット	抄録-DWPI 優位	1	2	3	4	5	6	7	8	9	10
自動分類前	レコード数	1	5	58	68	68	12	16	98	9	4
優先権主張年 (最先)	レコード数	1	5	58	68	68	12	16	98	9	4
1	72	C01:安全性	1	1	6	7	9	17	21	8	1
2	98	C02:制御			3	17	8	22	32	16	
3	63	C03:油圧			11	9	7	13	19	4	
4	39	C04:表示	1		4	6	6	6	11	5	
5	52	C05:作業効率			8	10	2	11	19	1	1

606件のうち、256件を教師データとして分類コードを付与

自動分類の結果をもとに再分類を行うことで、Unclassifiedを削減

リセット	優位性	1	2	3	4	5	6	7	8	9	10	11
自動分類後(1回目)	レコード数	1	5	58	68	68	12	16	98	9	4	
優先権主張年 (最先)	レコード数	1	5	58	68	68	12	16	98	9	4	
1	142	C01:安全性										
2	182	C02:制御										
3	79	C03:油圧										
4	82	C04:表示										
5	133	C05:作業効率										
6	100	Unclassified	1									

リセット	優位性 (インボ)	1	2	3	4	5	6	7	8	9	10	11
自動分類後(2回目)	レコード数	1	5	58	68	68	12	16	98	9	4	
優先権主張年 (最先)	レコード数	1	5	58	68	68	12	16	98	9	4	
1	149	C01:安全性										
2	211	C02:制御										
3	79	C03:油圧										
4	84	C04:表示										
5	152	C05:作業効率										
6	45	Unclassified	1									



## 自動分類結果による再分類 -手順-

- ① 1 回目自動分類後の結果をみてUnclassifiedに分類を付与
- ↓
- ② 自動分類結果グループをフィールド化
- ↓
- ③ 自動クラシファイア設定から 1 回目の自動分類の結果をエクスポート
- ↓
- ④ 手順 3 でエクスポートしたファイルをインポートし、分類カテゴリを作成
- ↓
- ⑤ 手順 4 でインポートして作成した分類カテゴリで再分類を行う
- ↓
- ⑥ 再度の自動分類を実施



# 1 1回目自動分類後の結果をみてUnclassifiedに分類を付与

※手順1は必須ではありません。より正確な分類やUnclassifiedを多く削減する場合は実施すると有効です。

## (手動分類付与方法①)

付与する際、フレーズ分解した文章（例では新規性:NLP/フレーズ）×自動分類した結果のマトリクスを作成し、Unclassifiedの案件をクリックし、タイトルウィンドウから該当案件を選択した状態で右クリックするとレコード分類の画面が表示され、キーワードを見ながら分類を検討できるので便利です。

	レコード数	出現回数	新規性	A01:シリンダー	A04:ジブ	A02:アウトリガ	A03:ブーム	A05:ディスプレイ	A06:ポンプ	A07:アーム	Unclassified
1	1	1	1 06d6w2j2o4w1 b6o439b4qd1 9363lx1 v6mn								
2	1	1	1 61 b1 d445a2h6c1 31 6m3m1 t2p554h6e2f4z 1u41								
3	1	1	1 614c3o2d433y51 6zm4b6o2v965y1 52y636n								
4	1	1	1 81 61 7435b1 u1 a3b3j2j6t6c3t724k4aa4w5v								
5	1	1	1 k6r2t3y3xa2d2e4z6h4s6c4h2s3r6u3t711 t 1s								
6	1	1	1 p273m5t1 x53921 m2e6z2v31 3u3j68323t2y 3t								
7	1	1	1 vy4y5r686xl p6n4b476635h241 32n3x523q 6i								
8	1	1	275m1 h1j3e1 r401 04l29303eyl 4 e44c 59								
9	1	1	2b4w1 71 5z5rf3a1 l1 134m4x2n1 y 2i94p								
10	1	1	2d2y4x6u221 d5f2a1 4h4n1 m5l1 13e2g734s3 5								
11	1	1	2h6661 3j48591 4i2x42o2q3f3x1 u31 28486n								

リセット		抄録-DWPI 新規性: NLP/フレーズ	1	2	3	4	5	6	7	8
		レコード数	97	77	98	99	10	85	29	41
新規性	レコード数	共起レコード数	A01:シリンダー	A02:アウトリガ	A03:ブーム	A04:ジブ	A05:ディスプレイ	A06:ポンプ	A07:アーム	Unclassified
1	7	operation amount			1		1	1		4
2	79	apparatus	10	9	41	1	18	12	4	4
3	76	state	15	11	30	10	16	9	2	4
4	52	operation	8	6	17	1	16	13		4
5	41	vehicle	7	7	17	1	4	5		4
6	27	engine	2	2			2	20	2	4
7	32	control unit	1	3	8	1	13	5	1	3
8	57	position	6	10	34	2	7	1	4	3
9	15	rotatably		1	11	1			1	3
10	21	structure	3	2	14	2			1	3

## タイトルウィンドウ

タイトルウィンドウ  
DWPI アクセション番号  
"4 タイトル, 1 選択"  
201650290M  
201723156Q  
201867735V  
201896741T

- レコードを表示
- 類似レコードを検索
- タイトル名をコピー
- すべて選択
- 選択したレコードをグループに追加
- 選択項目からサブデータセットを作成
- 新しいデータセットから選択したレコードを除外
- 新しいデータセットから選択したレコードを除外しない

レコードを分類

分類  
新規性

カテゴリ

- ☒ A01:シリンダー
- ☐ A04:ジブ
- ☐ A02:アウトリガ
- ☒ A03:ブーム
- ☐ A05:ディスプレイ
- ☐ A06:ポンプ
- ☐ A07:アーム
- ☐ Unclassified

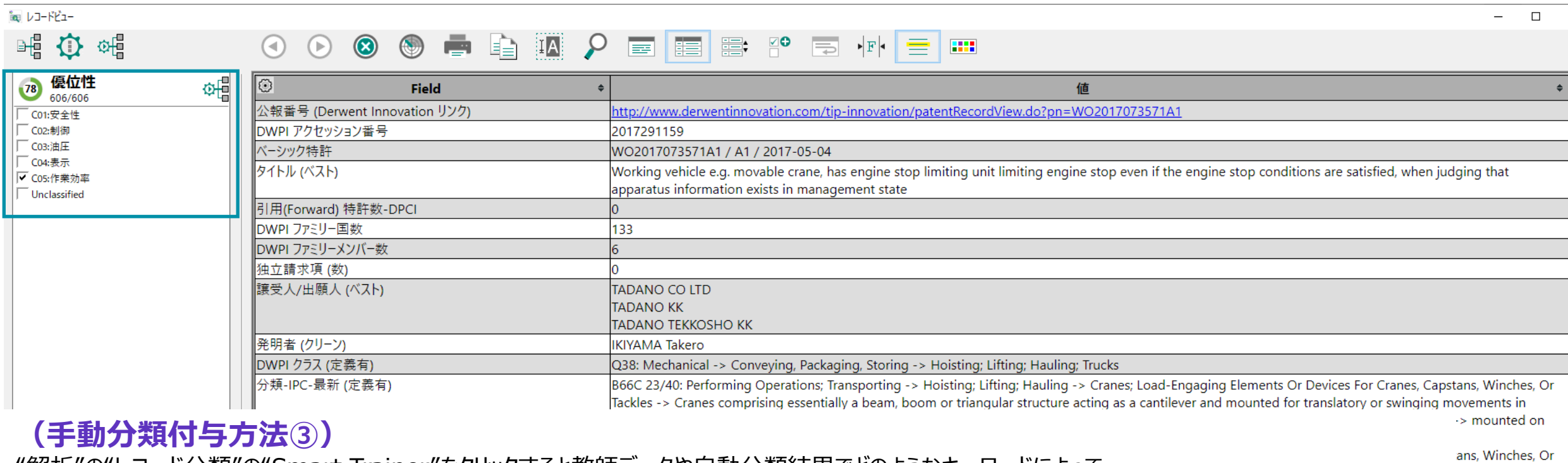
レコードを表示  
類似レコードを検索  
タイトル名をコピー  
すべて選択  
選択したレコードをグループに追加  
選択項目からサブデータセットを作成  
新しいデータセットから選択したレコードを除外  
新しいデータセットから選択したレコードを除外しない

レコードを分類 > 新規性

Unclassifiedのものに他の分類を付与します

## (手動分類付与方法②)

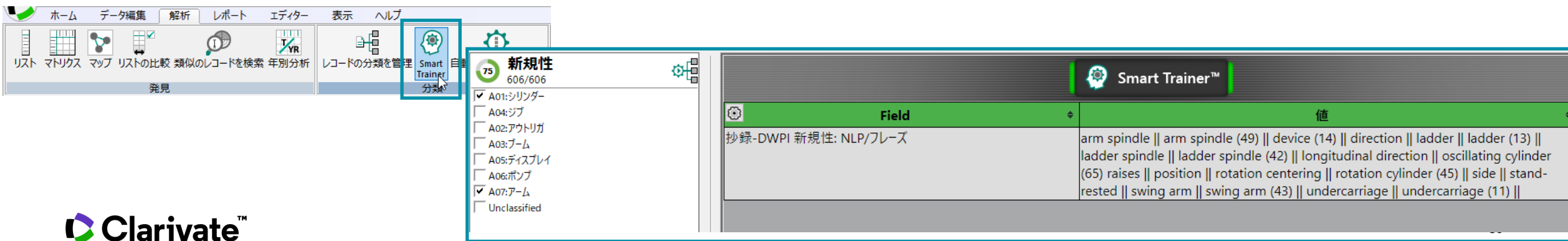
Unclassifiedのレコードビューを開き、各レコードの詳細内容を確認しながらレコードビュー左端にある分類を設定できます。



Field	値
公報番号 (Derwent Innovation リンク)	<a href="http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=WO2017073571A1">http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=WO2017073571A1</a>
DWPI アクセセッション番号	2017291159
ベーシック特許	WO2017073571A1 / A1 / 2017-05-04
タイトル (ベスト)	Working vehicle e.g. movable crane, has engine stop limiting unit limiting engine stop even if the engine stop conditions are satisfied, when judging that apparatus information exists in management state
引用(Forward) 特許数-DPCI	0
DWPI ファミリー国数	133
DWPI ファミリーメンバー数	6
独立請求項 (数)	0
譲受人/出願人 (ベスト)	TADANO CO LTD TADANO KK TADANO TEKKOSHO KK
発明者 (クリーン)	IKIYAMA Takero
DWPI クラス (定義有)	Q38: Mechanical -> Conveying, Packaging, Storing -> Hoisting; Lifting; Hauling; Trucks
分類-IPC-最新 (定義有)	B66C 23/40: Performing Operations; Transporting -> Hoisting; Lifting; Hauling -> Cranes; Load-Engaging Elements Or Devices For Cranes, Capstans, Winches, Or Tackles -> Cranes comprising essentially a beam, boom or triangular structure acting as a cantilever and mounted for translatory or swinging movements in ans, Winches, Or

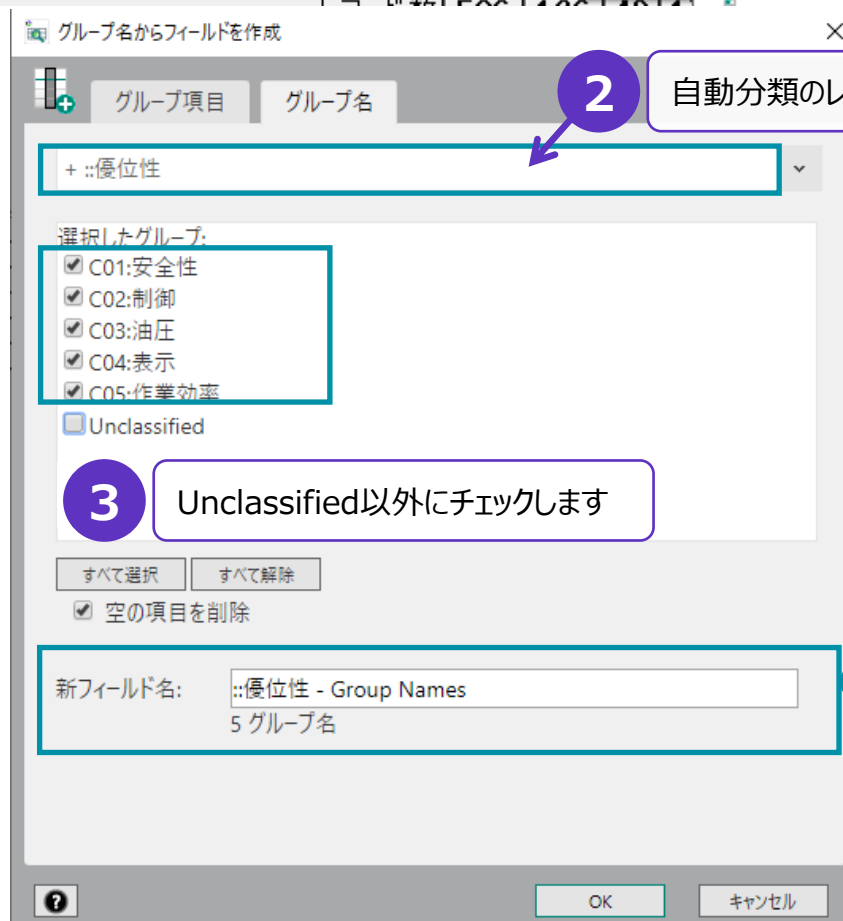
## (手動分類付与方法③)

“解析”の“レコード分類”の“Smart Trainer”をクリックすると教師データや自動分類結果でどのようなキーワードによって、分類カテゴリが付与されているのか参照しながら画面左端にある分類を設定できます。



Field	値
抄録-DWPI 新規性: NLP/フレーズ	arm spindle    arm spindle (49)    device (14)    direction    ladder    ladder (13)    ladder spindle    ladder spindle (42)    longitudinal direction    oscillating cylinder (65) raises    position    rotation centering    rotation cylinder (45)    side    stand-rested    swing arm    swing arm (43)    undercarriage    undercarriage (11)

## 2 自動分類結果グループをフィールド化

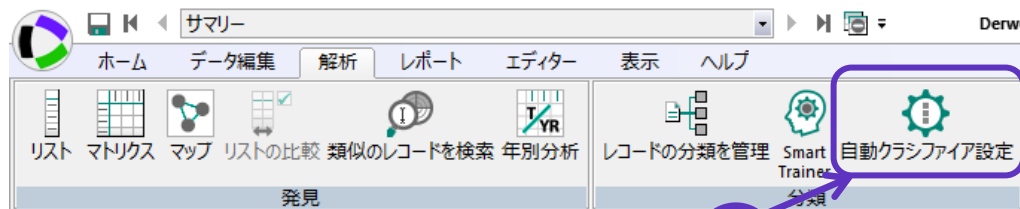


自動分類のレコード分類フィールドを選択します

1 回目の自動分類結果をもとにフィールドが作成されます

	レコード数	出現回数	::優位性 - Group Names
1	142	142	C01:安全性
2	182	182	C02:制御
3	79	79	C03:油圧
4	82	82	C04:表示
5	133	133	C05:作業効率

### 3 自動クラシファイア設定から 1 回目の自動分類の結果をエクスポート

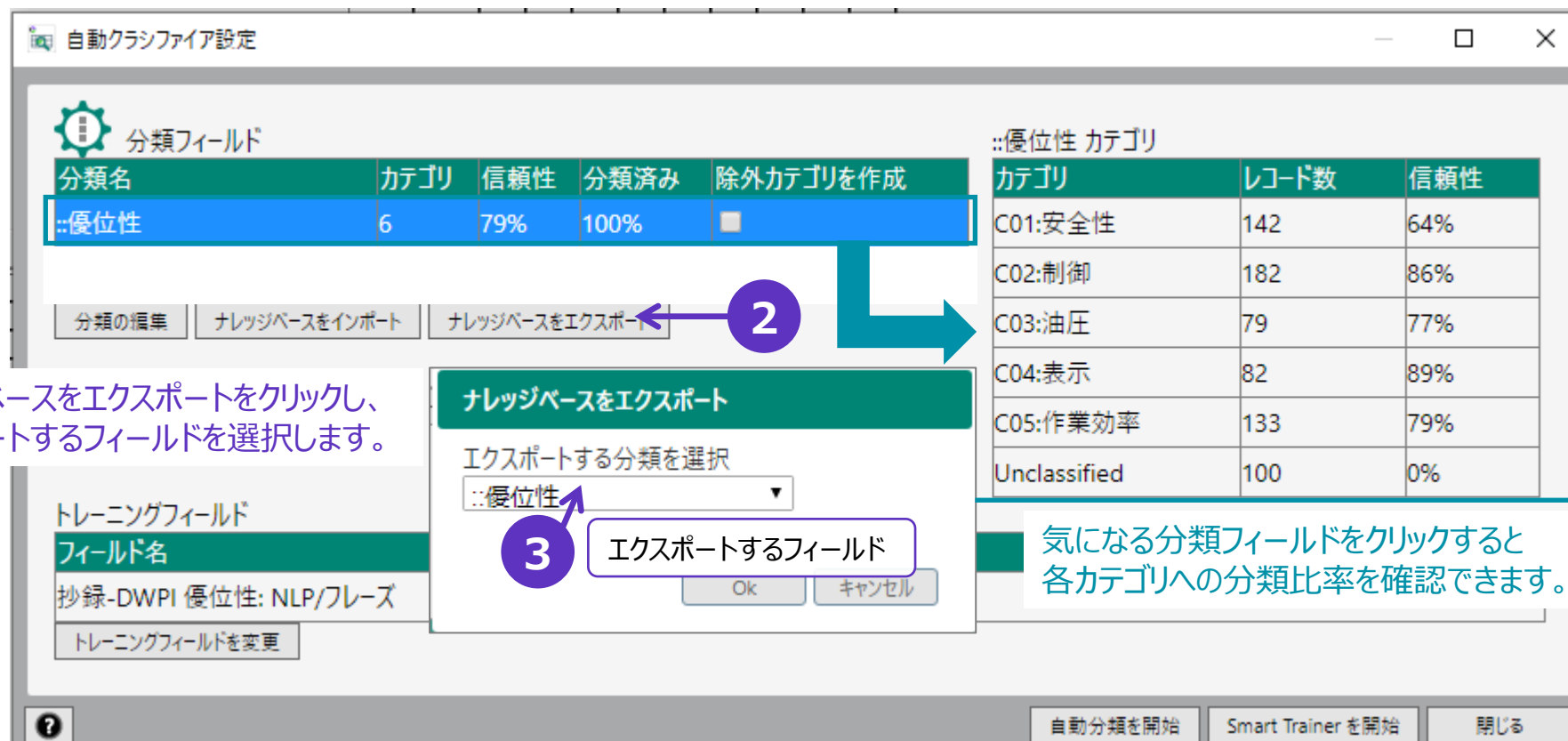


1

4

エクスポートしたデータを拡張子vpkbで保存します。

ファイル名(N): 優位性vpkb  
ファイルの種類(T): 自動クラシファイアのナレッジベース (\*.vpkb)  
保存(S) キャンセル



## 4 手順3でエクスポートしたファイルをインポートし、分類カテゴリを作成

自動クラシファイア設定

分類フィールド

分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::優位性	6	79%	100%	<input type="checkbox"/>

分類の編集

1 手順3でエクスポートしたファイルをインポートします。

トレーニングフィールド

フィールド名

抄録-DWPI 優位性: NLP/アレーズ

自動分類を開始 Smart Trainer を開始 閉じる

::優位性 カテゴリ

カテゴリ	レコード数	信頼性
C01:安全性	142	64%
C02:制御	182	86%
C03:油圧	79	77%
C04:表示	82	89%
C05:作業効率	133	79%
Unclassified	100	0%

インポートした分類フィールドは分類がされていないので分類0%、レコード数0となっています。

ファイル名(N): 優位性.vpkb

ファイルの種類(T): Knowledge Base Files (\*.vpkb)

☐ 読み取り専用ファイルとして開く(R)

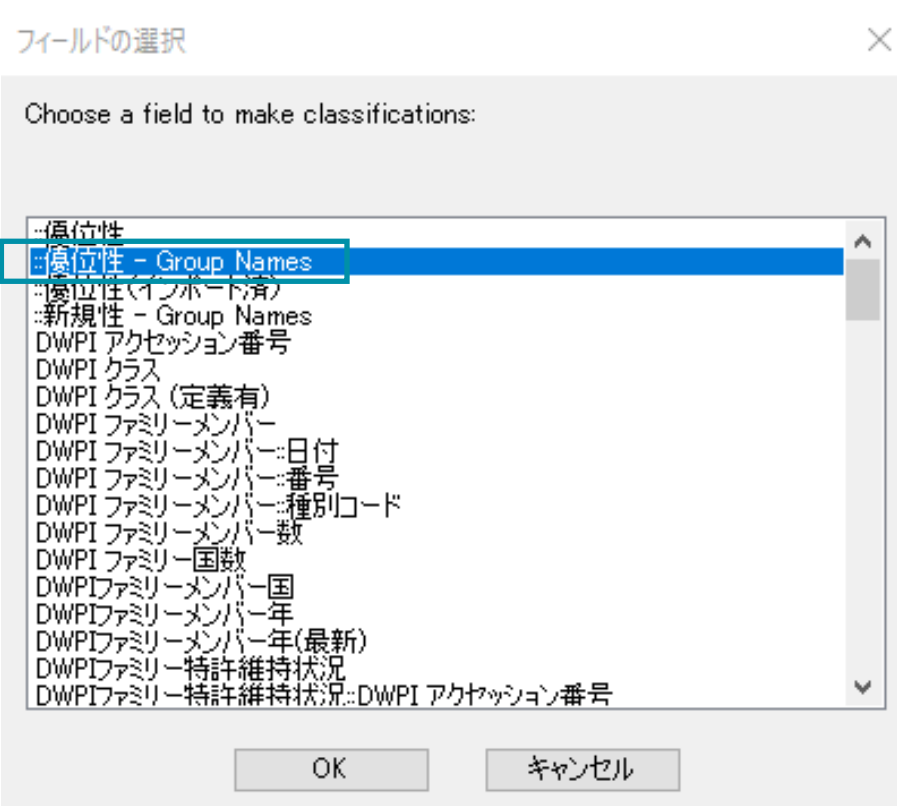
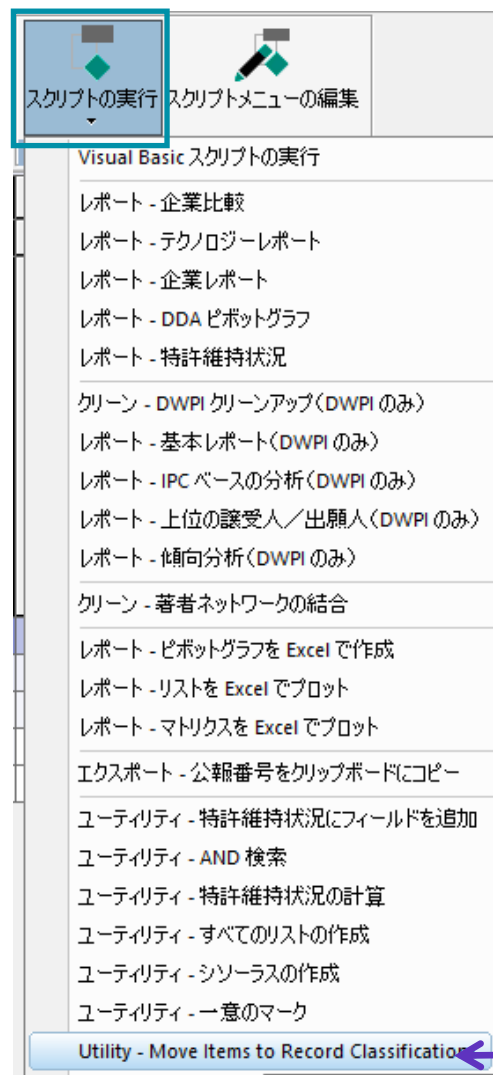
分類フィールド

分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::優位性	6	79%	100%	<input type="checkbox"/>
::優位性 (インポート済)	5	79%	0%	<input type="checkbox"/>

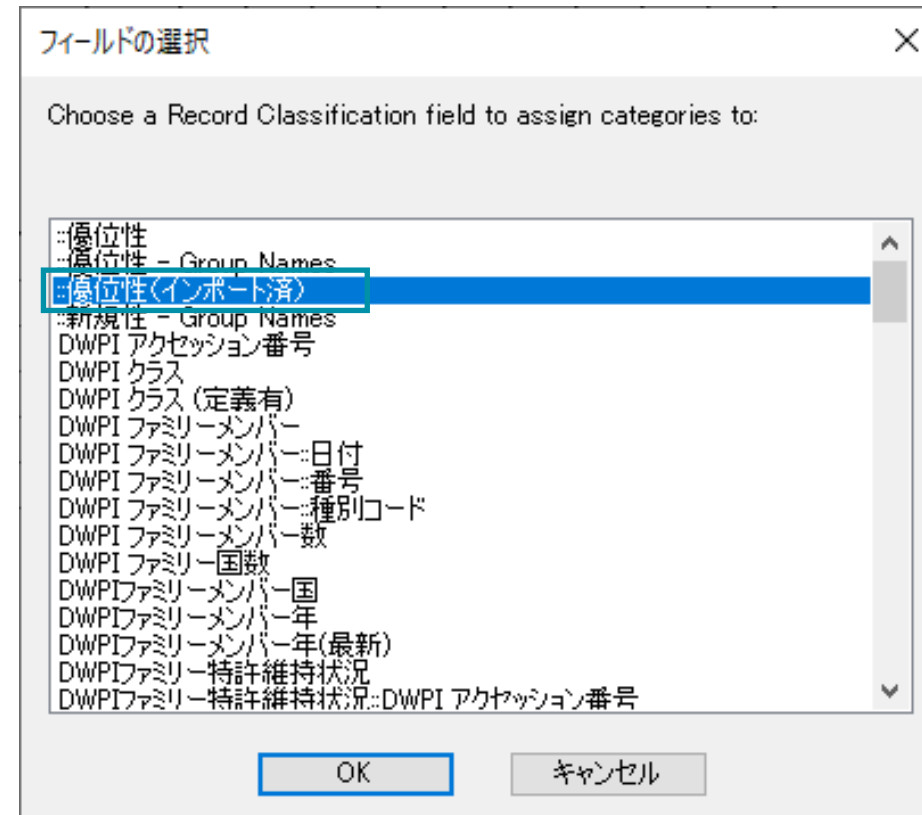
::優位性 (インポート済) カテゴリ

カテゴリ	レコード数	信頼性
C01:安全性	0	64%
C02:制御	0	86%
C03:油圧	0	77%
C04:表示	0	89%
C05:作業効率	0	79%

## 5 手順 4 でインポートして作成した分類カテゴリで再分類を行う

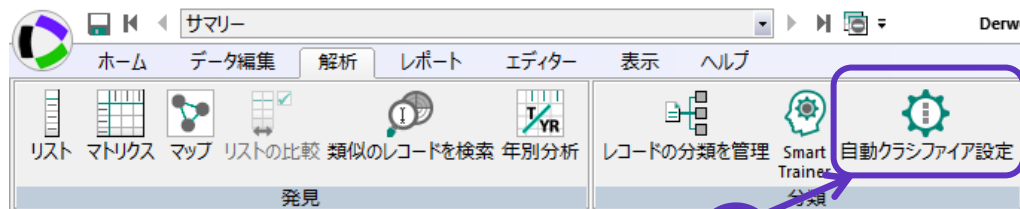


2 手順 2 で作成したフィールドを選択します。

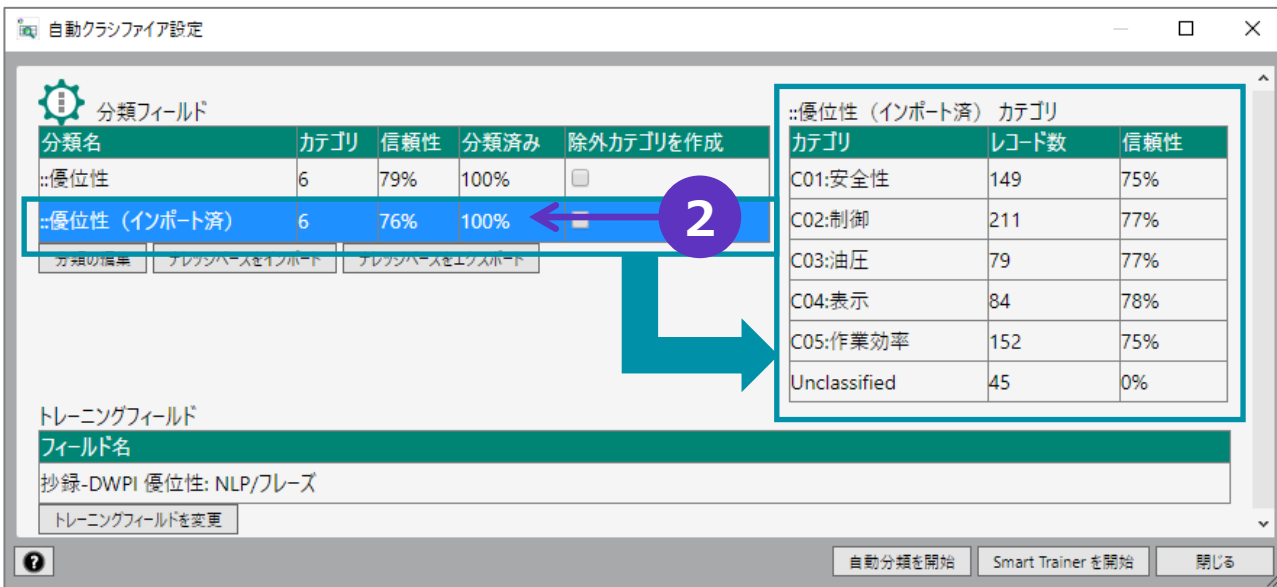


3 手順 4 でインポートした (インポート済) フィールドを選択します。

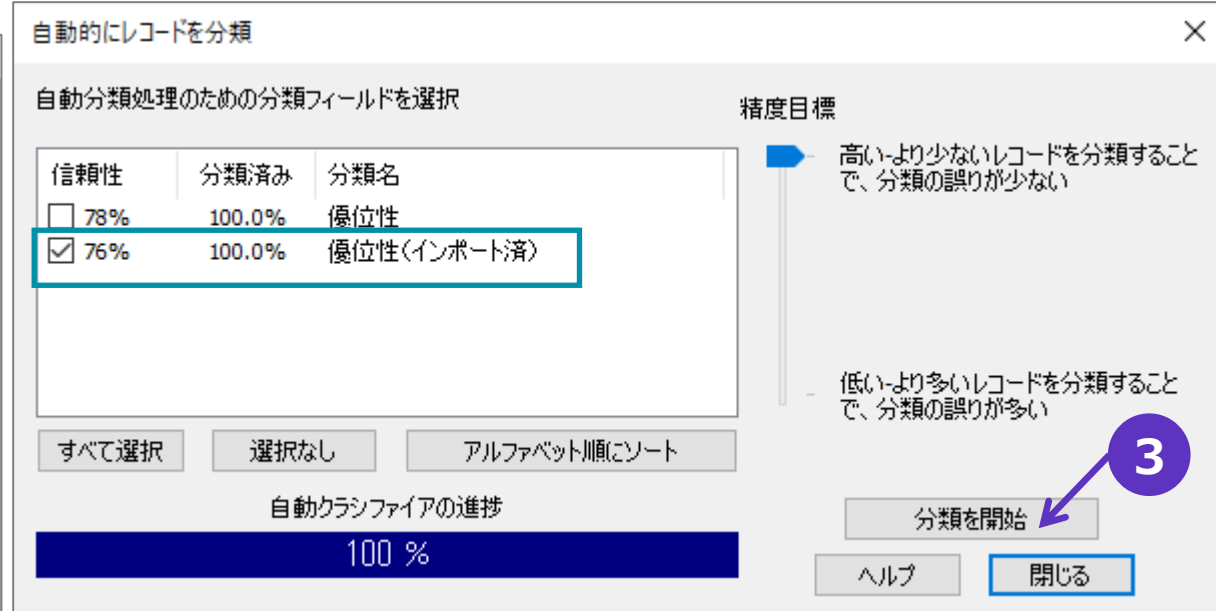
## 6 再度の自動分類を実施



自動分類後の手順 4 で作成した分類カテゴリへの分類結果を自動クラシフィア設定で確認することができます。



自動的にレコードの分類画面が開くので、手順 4 でインポートした分類カテゴリを選択し、精度目標を設定し、分類開始をクリックします。



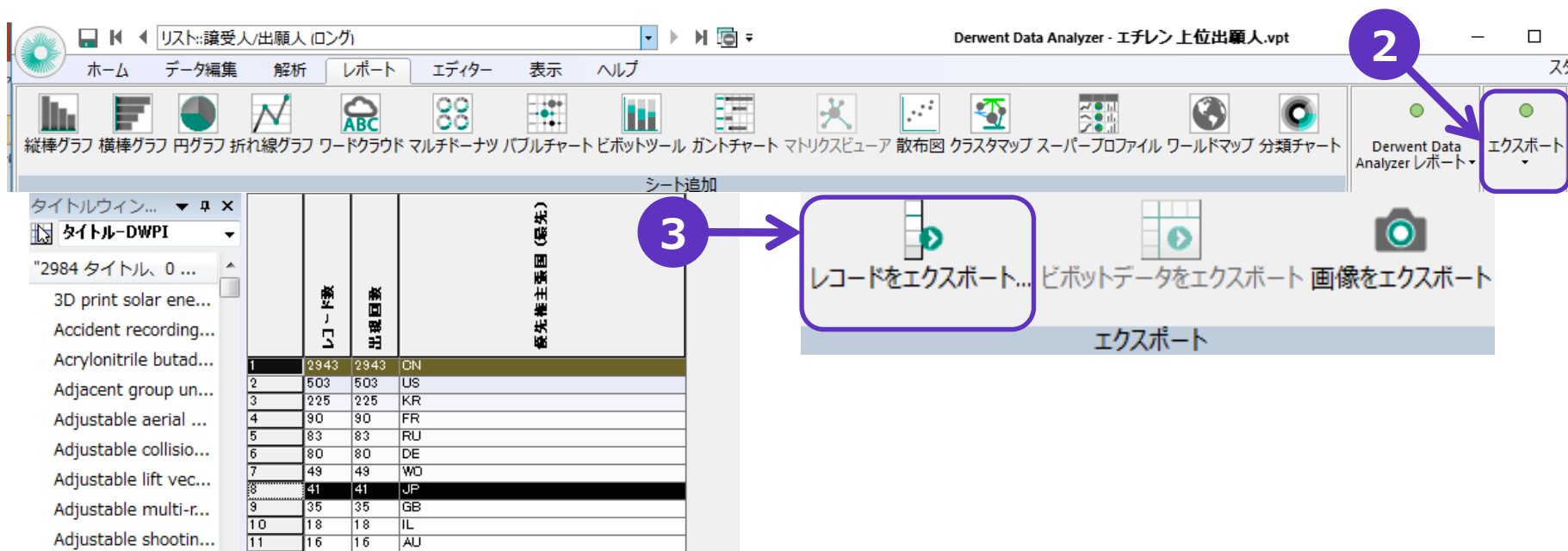
様々な処理：

- ・Excelへのデータ出力
- ・シソーラスの活用



# Excelへのデータ出力

DDA上で様々データ整理・編集、グループ分け等を行ったデータは、他でも利用できるよう、CSV形式でExcelに出力することができます。



1

出力したいデータの範囲がある場合、  
範囲を選択。  
全レコードの場合は、必要ありません。

## Excelへのデータ出力

出力したデータは、再度DDAで分析作業を行うために、DDAに取り込むことが可能です。CSVファイルをExcel形式で保存します。 ※自動分類（レコードの分類）を含んで出力する場合は、CSVファイルのみ対応しています。  
DDAにインポートする際は、File→Import Database Table(Excel, Access, etc)メニューを利用し、ExcelからのDDAインポートを行います。

**フィールドレコードをエクスポート**

レコードとフィールド | エクスポートするフィールドを変更

5 デフォルトの出力内容が表示されています。※変更する場合は[7]へ。

6 エクスポート: Excel

7 エクスポートするフィールドを変更

8 エクスポートする項目を選択  
左: エクスポート可能な項目  
右: エクスポート項目

9 順序を編集

10 エクスポート

自動分類  
結果の出力時  
はチェックON

ファイル形式を選択します。  
通常、カンマ区切りで問題ありません。

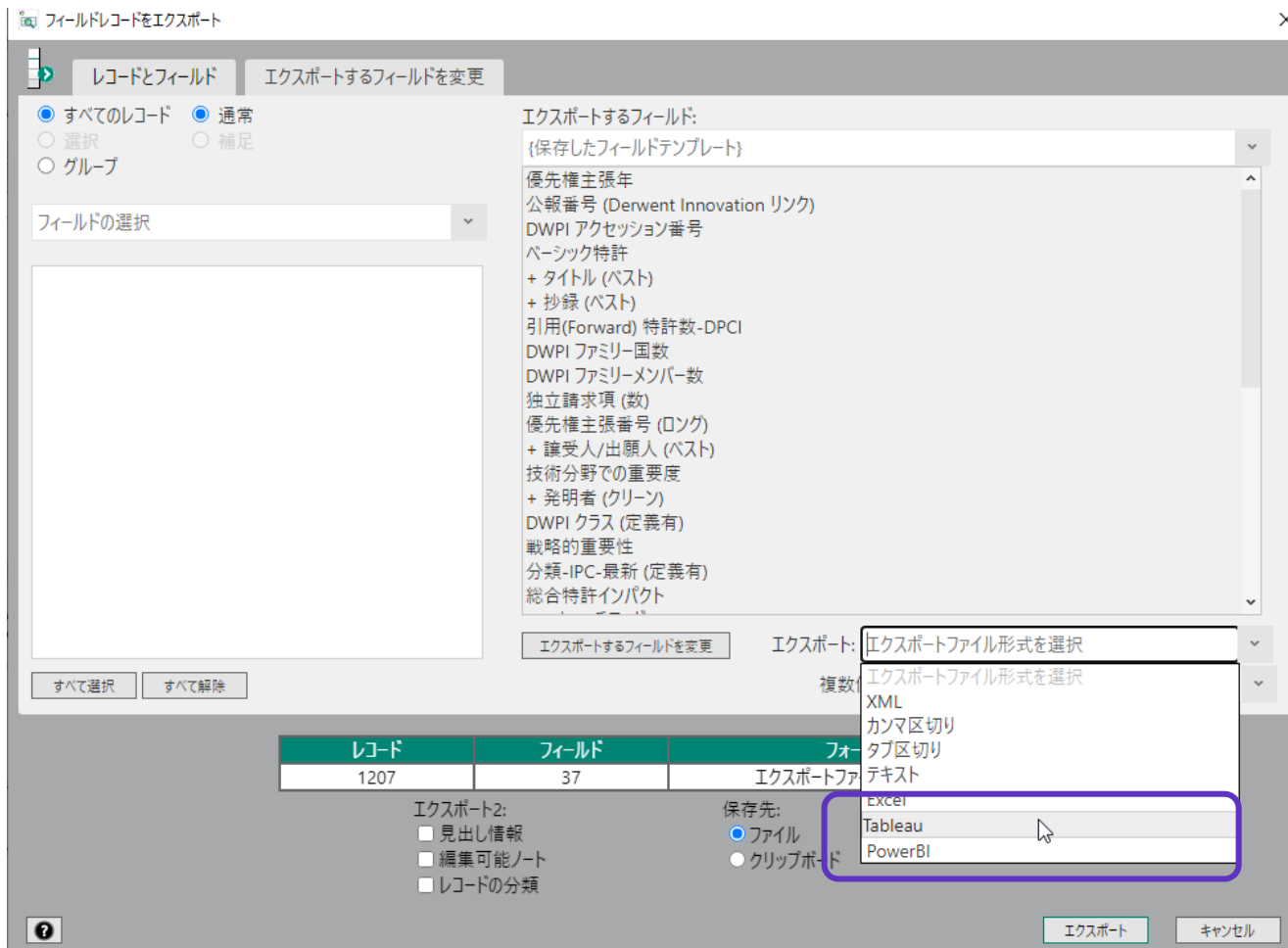
レコード	フィールド	フォーマット
63	25	Excel

エクスポート2:  
☐ 見出し情報  
☐ 編集可能ノート  
☒ レコードの分類

保存先:  
☐ ファイル  
☐ クリップボード

## エクスポートフォーマットTableauとPowerBIを追加

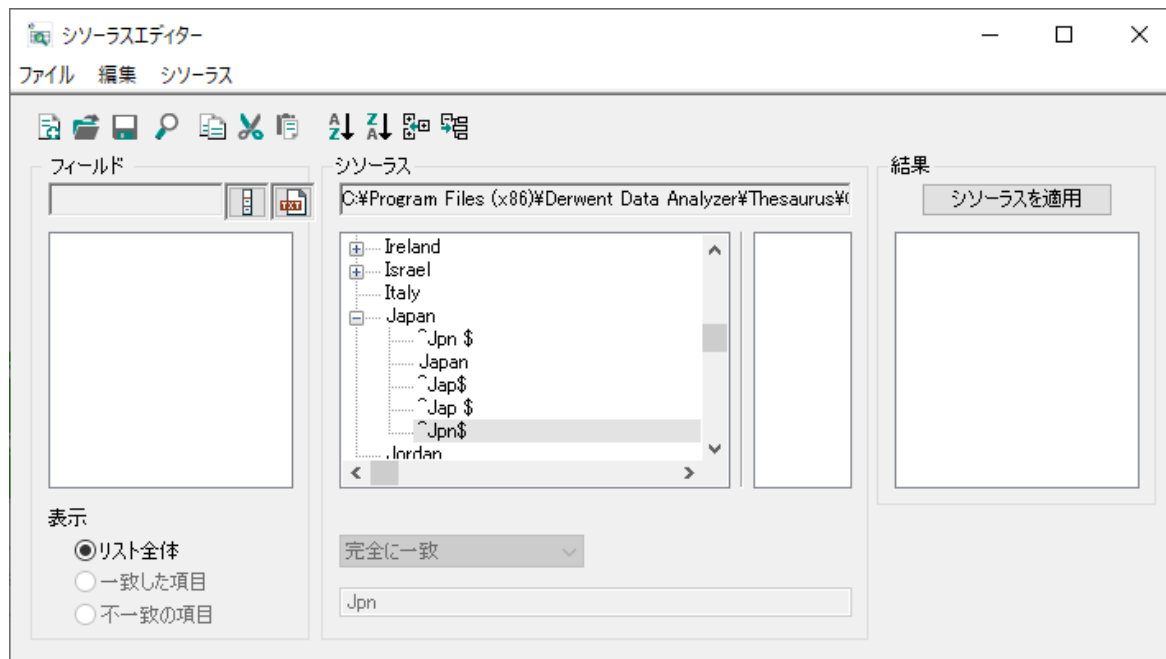
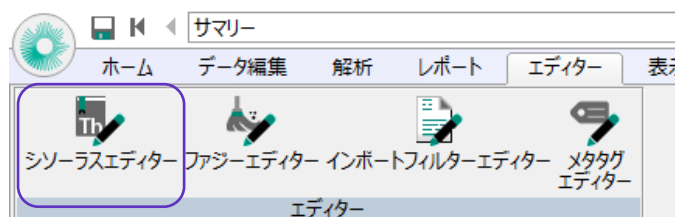
### ◆ Tableau や PowerBIへデータをインポートするためのエクスポートフォーマットを追加



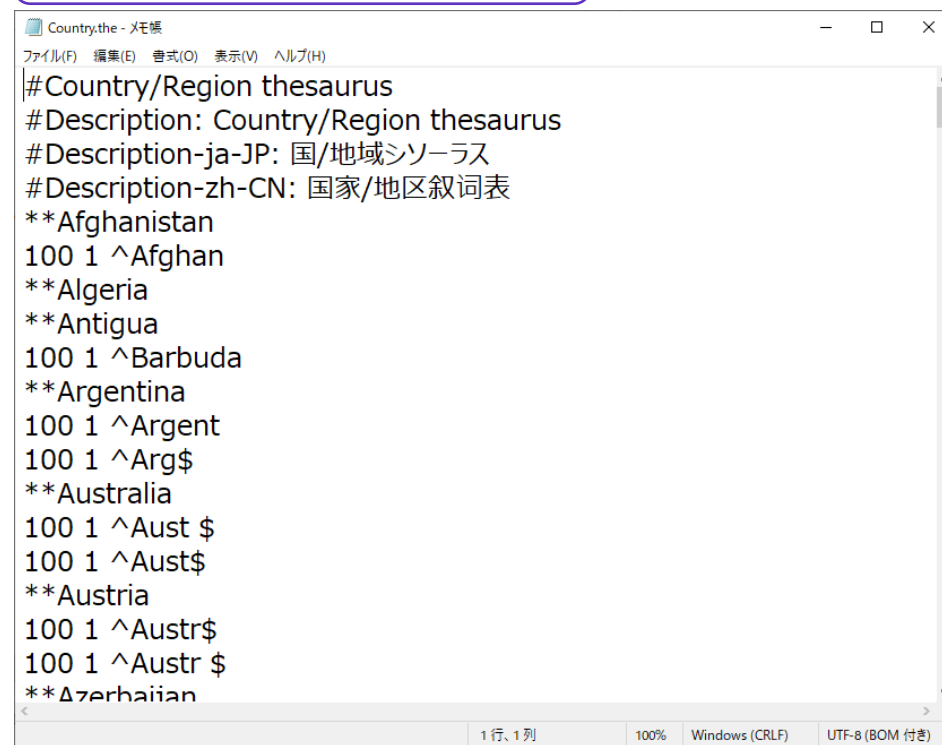
# シソーラスの活用

## シソーラスの独自編集

シソーラスエディタは、自分でシソーラスファイルを編集できる機能です。自分用にシソーラスを作成しておくと、次回以降の分析時に適用することで、文章・企業名などのデータ整理を簡単に素早く行うことが可能になります。シソーラスは、エディタを利用して編集することもできますし、テキストファイルでファイルを開いて編集することもできます。

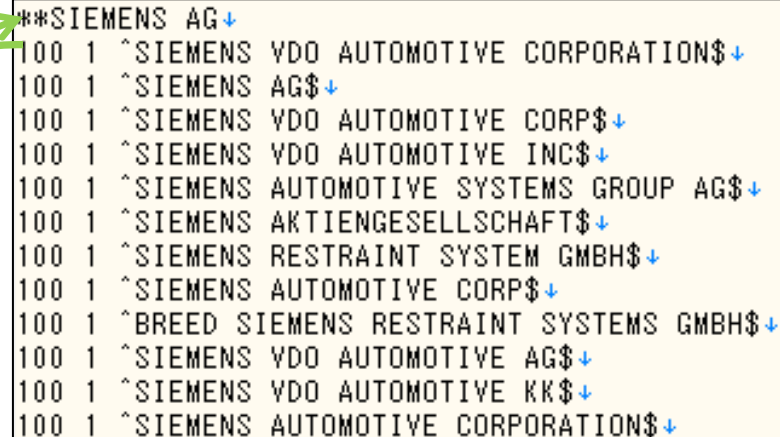


### テキストエディタでのシソーラスの編集



## シソーラスの活用 シソーラスの独自編集

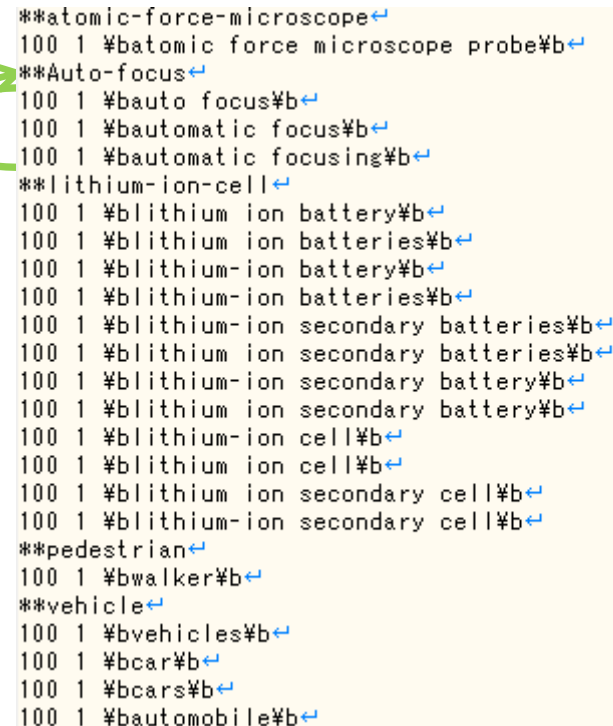
クリーニング作業で保存したシソーラスは、テキストエディタで開いて編集することができます。



```
##SIEMENS AG↓
100 1 ^SIEMENS VDO AUTOMOTIVE CORPORATION$↓
100 1 ^SIEMENS AG$↓
100 1 ^SIEMENS VDO AUTOMOTIVE CORP$↓
100 1 ^SIEMENS VDO AUTOMOTIVE INC$↓
100 1 ^SIEMENS AUTOMOTIVE SYSTEMS GROUP AG$↓
100 1 ^SIEMENS AKTIENGESELLSCHAFT$↓
100 1 ^SIEMENS RESTRAINT SYSTEM GMBH$↓
100 1 ^SIEMENS AUTOMOTIVE CORP$↓
100 1 ^BREED SIEMENS RESTRAINT SYSTEMS GMBH$↓
100 1 ^SIEMENS VDO AUTOMOTIVE AG$↓
100 1 ^SIEMENS VDO AUTOMOTIVE KK$↓
100 1 ^SIEMENS AUTOMOTIVE CORPORATION$↓
```

^ = 単語の最初がこの文字から始まる  
という指定

\$ = 単語の最後がこの文字で終わる  
という指定



```
##atomic-force-microscope↵
100 1 %batomic force microscope probe%b↵
##Auto-focus↵
100 1 %bauto focus%b↵
100 1 %bautomatic focus%b↵
100 1 %bautomatic focusing%b↵
##lithium-ion-cell↵
100 1 %blithium ion battery%b↵
100 1 %blithium ion batteries%b↵
100 1 %blithium-ion battery%b↵
100 1 %blithium-ion batteries%b↵
100 1 %blithium-ion secondary batteries%b↵
100 1 %blithium ion secondary batteries%b↵
100 1 %blithium-ion secondary battery%b↵
100 1 %blithium ion secondary battery%b↵
100 1 %blithium-ion cell%b↵
100 1 %blithium ion cell%b↵
100 1 %blithium ion secondary cell%b↵
100 1 %blithium-ion secondary cell%b↵
##pedestrian↵
100 1 %bwalker%b↵
##vehicle↵
100 1 %bvehicles%b↵
100 1 %bcar%b↵
100 1 %bcars%b↵
100 1 %bautomobile%b↵
```

## 文字列処理を行いやすくするシソーラスの活用

出願人名の処理や文字列の処理を行う際、-（ハイフン）の有無や語尾の違いなど、様々なバラつきがあり、まとめることが難しい場合があります。

シソーラスを作成して、検索と置換を利用することにより、統一表記に書き換えてしまうことができるので、そのバラつきを統制することができます。その後で、クリーニング等の処理を行うことにより、素早く正確に適切な単語の抽出を行うことができます。

【データ編集】→【シソーラス】→【検索と置換】

6	2	5	Atomic force microscope for observing bio-molecules, has phase d
7	2	5	Atomic force microscope probe for use in e.g. micro electronic dev
8	2	4	Atomic force microscope tweezers for use in scanning probe micr
9	2	2	Auto-focus microscope for inspecting specimen has decision unit v
10	2	2	Automatic focus adjustment in microscope computes focussing range
11	2	2	Automatic focusing detector for microscope used in area that is d



6	2	5	atomic-force-microscope for observing bio-molecules, has phase
7	2	5	atomic-force-microscope probe for use in e.g. micro electronic d
8	2	4	atomic-force-microscope tweezers for use in scanning probe mic
9	2	2	Auto-focus adjustment in microscope computes focussing range
10	2	2	Auto-focus detector for microscope used in area that is divided
11	2	2	Auto-focus microscope for inspecting specimen has decision unit

【作成したシソーラス例】

\*置換.the - メモ帳

ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)

#Find Replace

\*\*Auto-focus

100 1 ^auto-focus

100 1 ^auto focus

100 1 ^automatic focus



## サービス全般に関するお問い合わせ

Tel:03-4589-3101

〒107-6118 東京都港区赤坂5丁目2番20号  
赤坂パークビル18階

## カスタマーサービス（ヘルプデスク）

Tel (フリーコール) : 0800-170-5577

Tel : 03-4589-3107

Email: [ts.support.jp@clarivate.com](mailto:ts.support.jp@clarivate.com)

サービス時間：月～金（祝祭日を除く）  
午前9時30分～午後5時30分

