



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

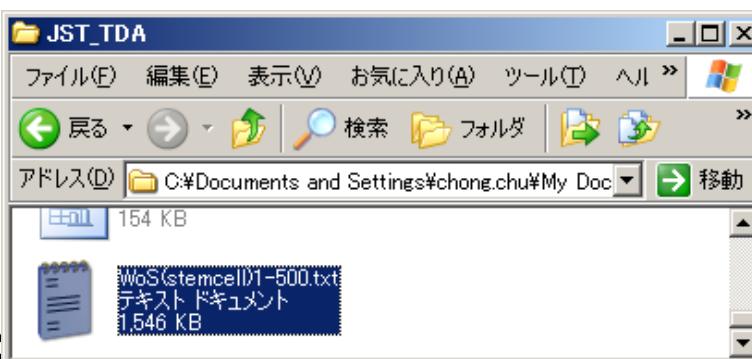
The screenshot shows a search result for 'Galicia and the Bay of Biscay: Lysosomal responses' with 162 records found. The 'Record Output' section is highlighted, showing the following steps:

- Step 1: ページ上で選択したレコード (Selected)
- Step 2: 著者名、タイトル、ジャーナル名 (Selected)
- Step 3: 文献管理ソフトにエクスポートするには? (Selected)

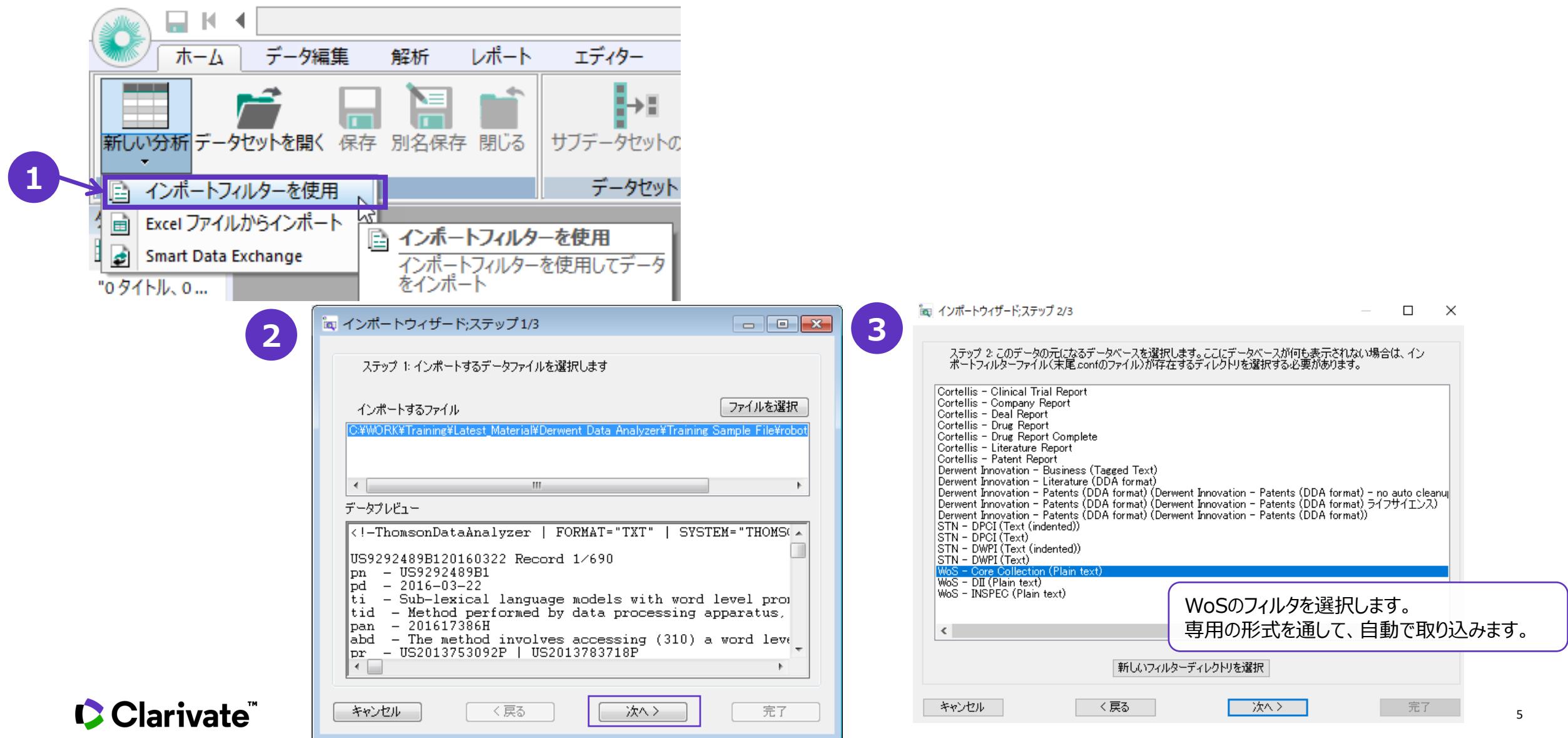
Under Step 3, the 'Save As' dropdown menu is open, showing various options:

- EndNote Web (selected)
- EndNote (disabled)
- ResearcherID (disabled)
- 文獻管理ソフト向けに保存 (disabled)
- BiTeXで保存 (disabled)
- HTMLで保存 (disabled)
- テキストで保存 (selected)
- タブ区切りで保存 (Win) (disabled)
- タブ区切りで保存 (Mac) (disabled)

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

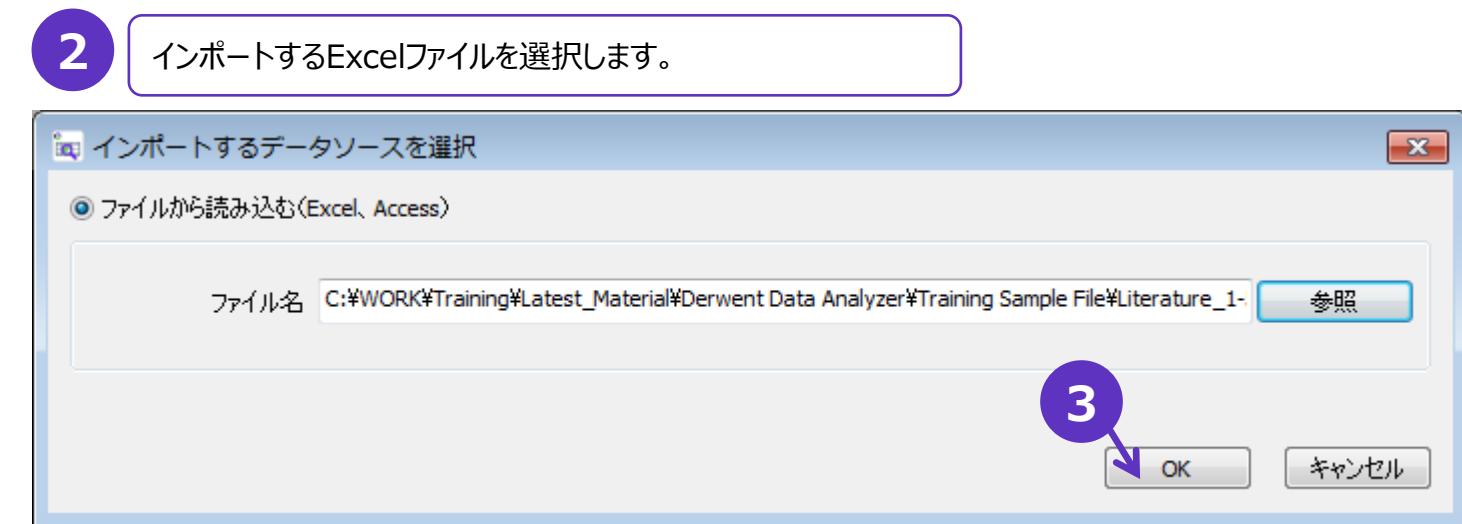
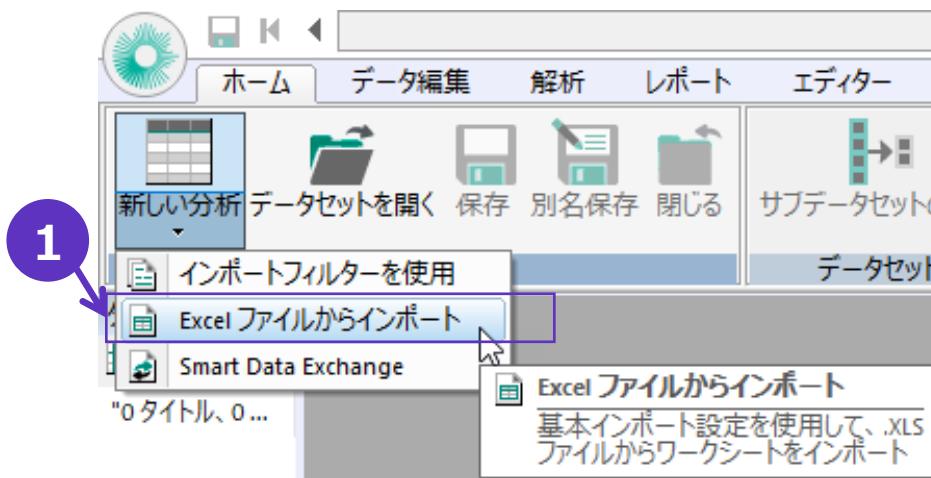
下の例のように、最初の行に各列名があり、2行目以降に各レコードの該当のフィールド情報が入った形式ファイルであれば利用可能です。

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できます。

【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 _A ₅ receptor - useful for the treatment of diminished	Alzheimers disease; Bleeding; Brain hemorrhage; Brain injury; Cerebral hypoxia; Cerebrovascular ischemia; Epilepsy; Huntingtons chorea; Hypoxia; Ischemia; Major depressive disorder; Parkinsons disease; Psychotic disorder				
WO-2006032273	30-Mar-2006	H Lundbeck A/S	New 2-acylaminothiazole derivatives are adenosine A _{2A} receptor modulators - useful for the treatment of neurological and psychiatric disorders.	Adenosine A _{2a} 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 β-secretase activity - useful for the treatment of Alzheimers disease.	Alzheimers disease	Beta secretase inhibitor; Neuroprotectant			Product
WO-2006034089	30-Mar-2006	Targacept Inc	New azaspiroalkane and azaspiroalkane compounds are nicotinic cholinergic receptor modulators - useful for the treatment of CNS disorders such as pre-senile dementia, Alzheimers disease, HIV dementia and Parkinsons disease.	Alzheimers 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; Alzheimers 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 CoMentis Inc; Zapaq Inc; University of Illinois; Oklahoma Medical Research Foundation	New benzisoxazole derivatives are PPAR _α 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		New bicyclic compounds are inhibitors of memapsin-2 β-secretase activity - useful for the treatment of Alzheimers	Alzheimers 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.	Alzheimers 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.	Alzheimers 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



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

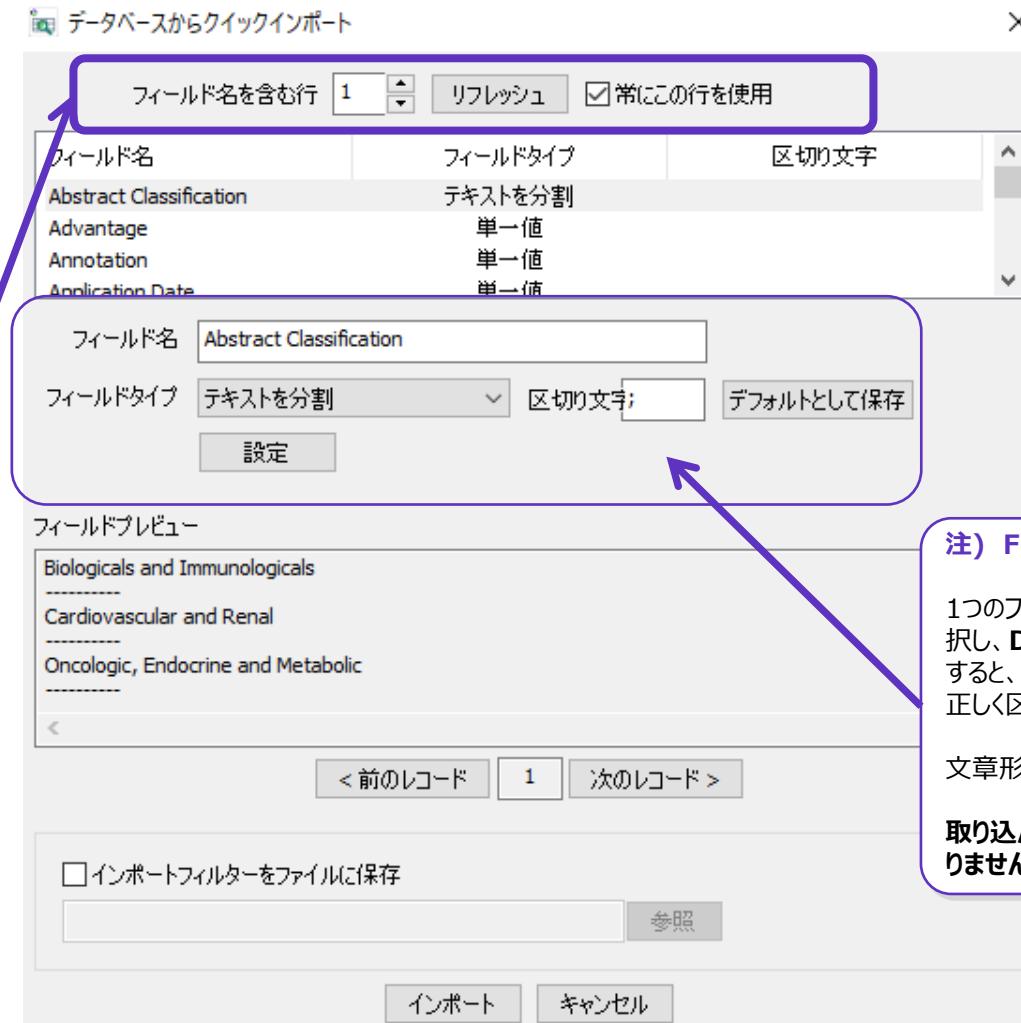
3

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。



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】

The screenshot shows the Derwent Data Analyzer (DDA1) application window. The menu bar includes リスト:Application Date, ホーム, データ編集 (selected), 解析, レポート, エディター, 表示, ヘルプ. The toolbar has icons for データセット (Subdataset creation, Record fusion, Duplicate record removal, Duplicate record combination), DWPIクリーンアップ (DWPI Clean Up), フィールド (Field management, Field merge, Add field import), and リストの比較 (List comparison). A context menu is open over a list of items, with '区切る' (Separate) highlighted. Examples are provided for each function:

- 【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(PIPE) 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：読み取り開始 - 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】

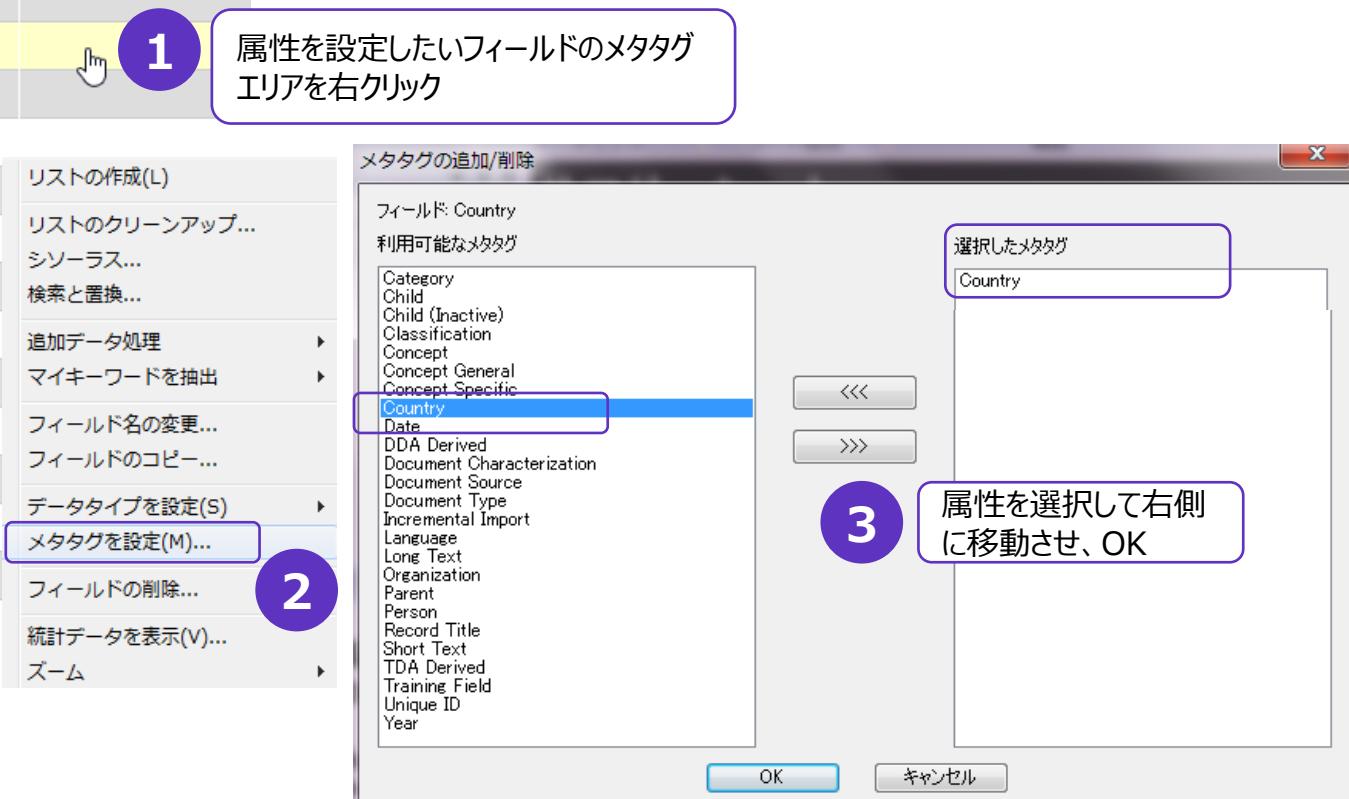
例：country情報（国コード）をワールドマップの軸に設定したい場合

サマリーシート						
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%			

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

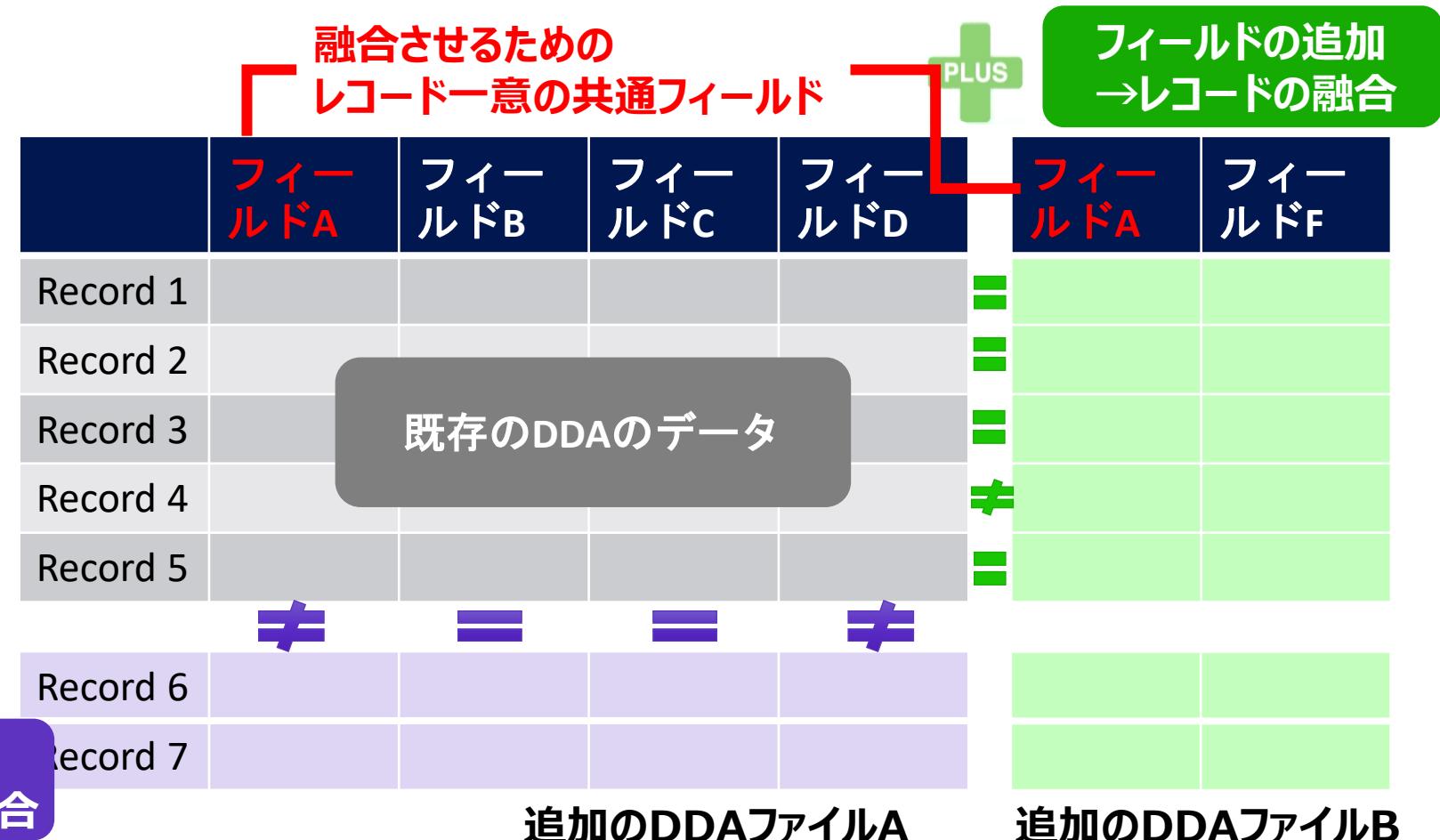


データの追加：複数ファイルの統合

- ・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) interface. The main window displays a summary sheet with 398 records. A callout bubble labeled '1' points to the bottom of the window where it says '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). Another callout bubble labeled '2' points to the 'Merge' button in the toolbar, which is highlighted with a blue circle.

Derwent Data Analyzer - DDA4

サマリー

ホーム データ編集 解析 レポート エディター 表示 ヘルプ

重複レコードの削除 重複レコードの結合 DWPI クリーンアップ リストのクリーンアップ シソーラス グループからフィールドを作成 フィールド フィールドの管理 フィールドのマージ 追加のフィールドをインポート グループの編集 リストの比較 グループヒソーラス グループの移動 グループ

サブデータセットの作成 引数データ データセットの融合 テーブル

タイトルウインドウ 詳細ウインドウ

サマリーシート レコード数: 398 列表示 i

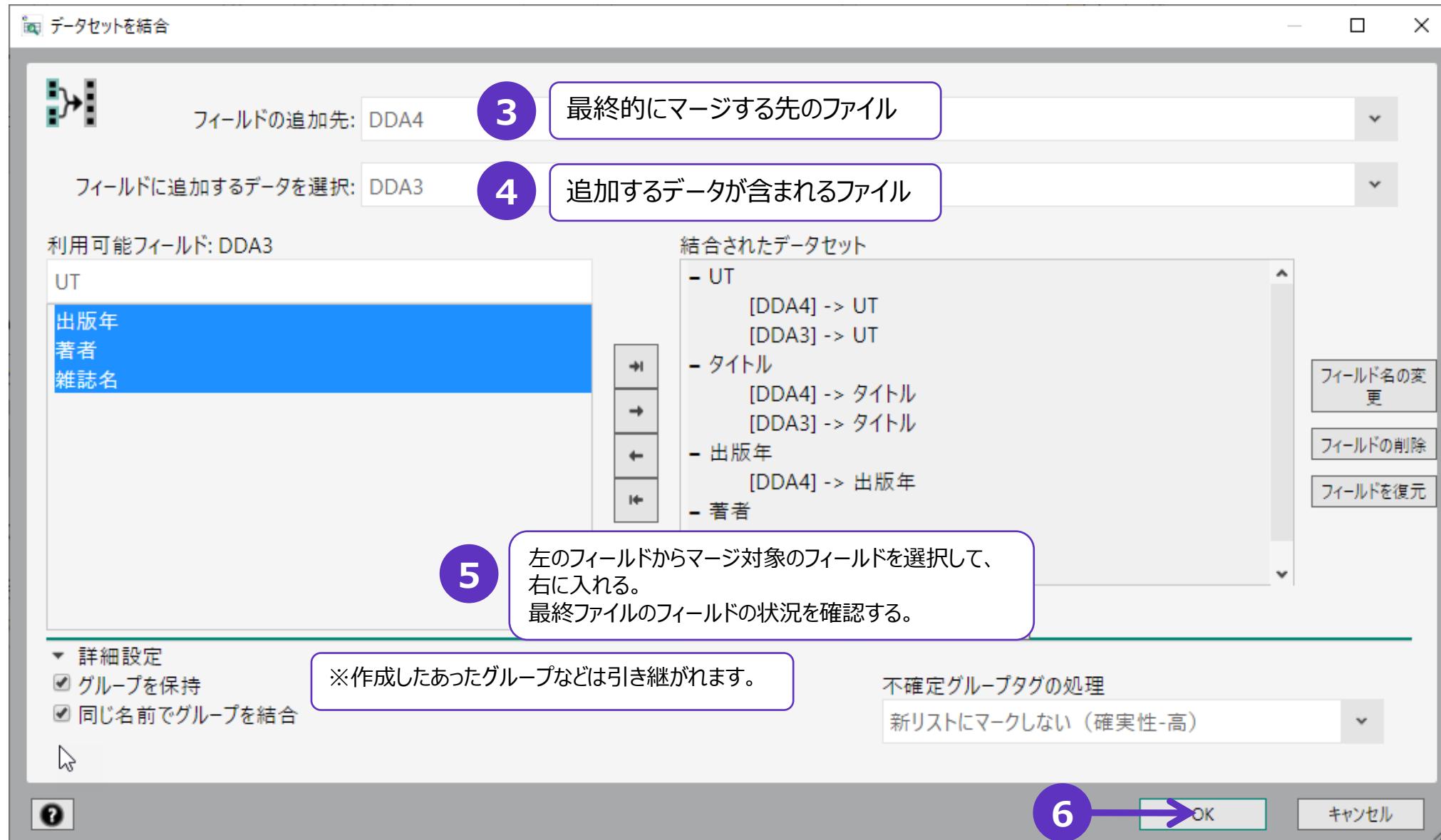
フィールド 項目数 グループ数 カバー率(%) データタイプ メタタグ

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%		

非表示のフィールド

サマリ DDA3 DDA4

後からデータを追加する（データセットの融合）



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、現在のフィールドではない、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 interface with the title "Derwent Data Analyzer - 追加フィールド.vpt". The ribbon menu is visible with tabs like Home, Data Edit, Analysis, Report, Editor, View, and Help. A callout bubble labeled "1" points to the bottom-left corner of the main data grid, which displays a summary of fields: フィールド (Field), 項目数 (Count), グループ数 (Group Count), カバー率 (%) (Coverage Rate), データタイプ (Data Type), and メタタグ (Metatags). The coverage rate for the last row (UT) is 100%. A callout bubble labeled "2" points to the "Merge Fields" button in the toolbar, which is highlighted with a purple arrow. The status bar at the bottom shows "マスターファイル.vpt" and "追加フィールド.vpt".

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。

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

3 最終的にマージする先のファイル

4 追加するフィールドが含まれるファイル

5 共通項フィールドは、
文字列で完全一致と設定する

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。

7 出来上がりのファイルの
フィールド状態の一覧

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
- + タイトル
- + 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

すべてのフィールドを開く

すべてのフィールドを閉じる

OK キャンセル

?

16

The screenshot shows the 'Merge from Existing Dataset' dialog box. Step 3 highlights the target file 'マスター・ファイル.vpt'. Step 4 highlights the source file '追加・フィールド.vpt'. Step 5 highlights the matching field setup where both 'UT' fields are set to '完全一致' (Exact Match). Step 6 shows the mapping of fields from the left list to the right list. Step 7 shows the resulting field status list on the right side of the dialog.

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

		DWPI Manual Codes																						
		# Records																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
		56	19	18	14	13	11	97	96	96	94	92	86	86	86	82	81	80	80	74	70	62		
		Show Values >= 1																						
		Coccourrence																						
		# of Records																						
		Alzheimer's disease	Parkinson's disease	Cancer	Neurodegenerative disease	Inflammatory disease	Cerebrovascular ischemia	Major depressive disorder	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	50	55	55	49	36	47	40	30	29	49	28	56	42	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-S1 6	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-E1 2	144	58	44	44	34	35	58	90	44	37	12	34	45	32	7	45	11	33	16	10	20	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	48	27	20	26	41	67	39	25	27	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	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

		DWPI Manual Codes																					
		# Records																					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		50	38	37	36	33	31	28	17	17	15	13	12	12	11	11	11	10	10	10	10	10	10
		Show Values >= 1																					
		Coccourrence																					
		# of Records																					
		Actions																					
		# Records																					
		Neuroprotectant																					
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	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				
8	16	B14-L06	16	6	13	11	8	7	7	7	4	6	2	6	5	3	3	2	4	2	6	3	2
9	16	B14-S01	17	4	8	4	11	6	16	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	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-S1 6	7	2	3	3	6	10	1	1	1	6	6	2	5		3	2	1				
13	14	B14-C01	5	2	7		7							7		4	1						
14	14	B14-E1 2	3	2	6		7	1	2					1	3	8			1		4		
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	6	9	1				1	13	8	1	1	3	1		1	
20	12	B14-N1 7	2	3		8	1	14	1					1	7		3	3	1	4		3	
21	12	B14-N03	4	5	4	8	2	9						3	3	9	6	2	5		1	1	
22	12	B14-J05																					

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 223	グループ数	カバー率 (%) 93%	データタイプ	現 メタタグ DDA 由来
(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, 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	

1

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。

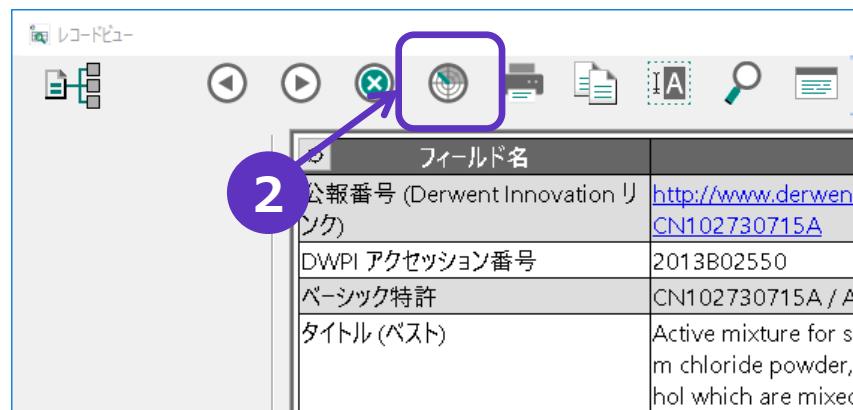
フィールドで、右クリックして、「メタタグを設定」をクリックすると、設定できます。

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。



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

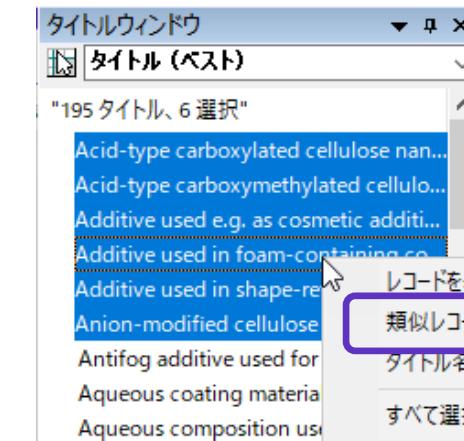
類似レコードを検索するときに表示しない

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

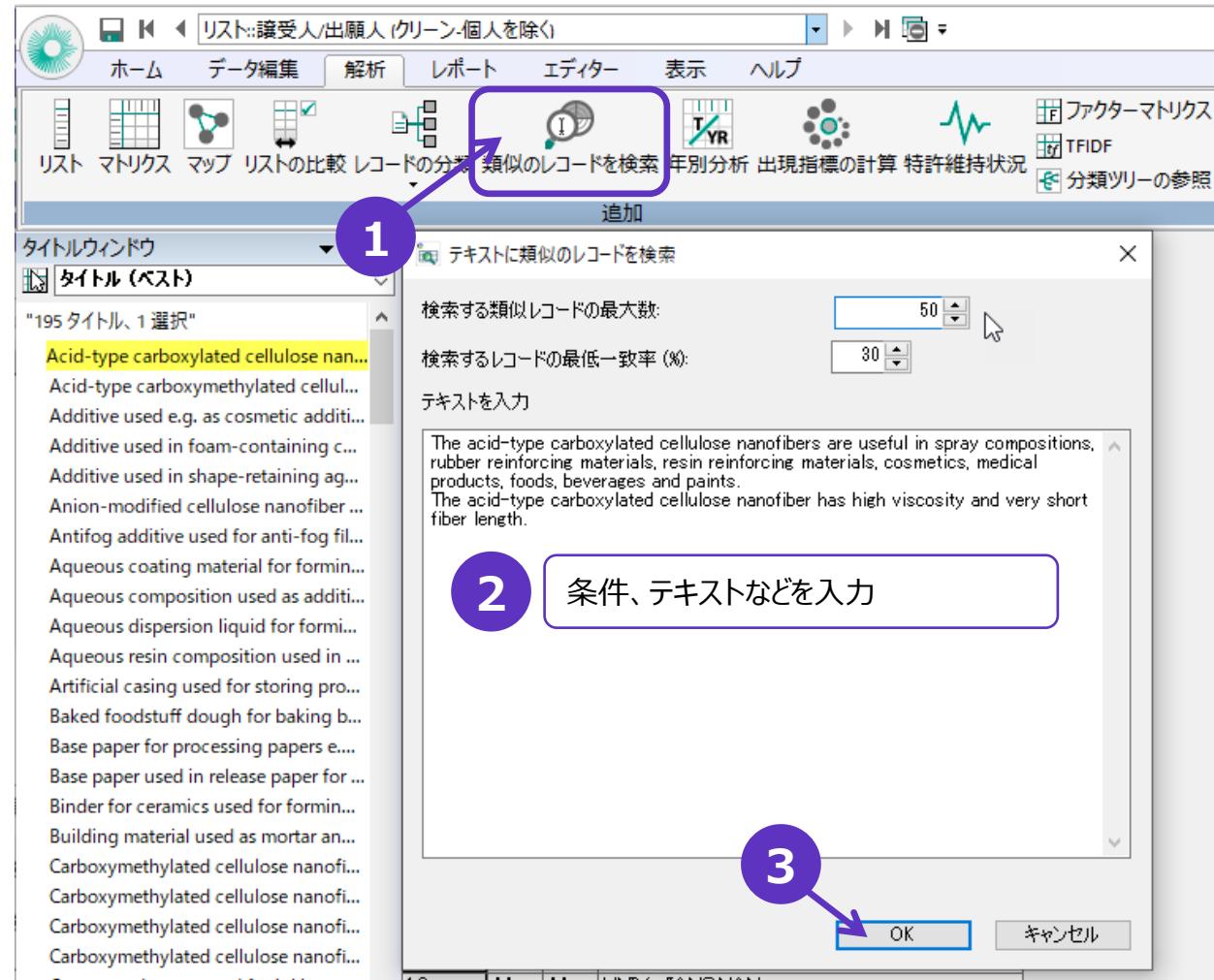
右クリックして「類似レコードを検索」をクリック。

The screenshot shows the 'Similar Record Display' window with 10 results (28% similarity). It lists various patent fields and their values, such as 'Publication number (Derwent Innovation link)', 'DWPI Accession number', 'Priority patent', and 'Title (Best)'. A purple box at the bottom right contains a note about the display format.

Field	値
公報番号 (Derwent Innovation リンク)	http://www.derwentinnovation.com/tip-innovation/patentRecordView.do?pn=US20160200950A1
DWPI アクセッション番号	2016427248
ペーシック特許	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



Field	値
公報番号 (Derwent Innovation リンク)	http://www.derwentinnovation.com/tip-innovation/patentRecordView. do?pn=WO2018116661A1
DWPI アクセッション番号	201850977D
ベース特許	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

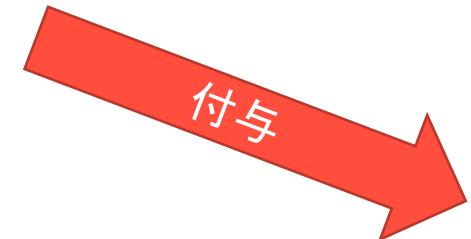
類似のレコードがパーセンテージとともに、レコード画面の形式で表示されます。

機械学習機能を利用して自動分類を行う

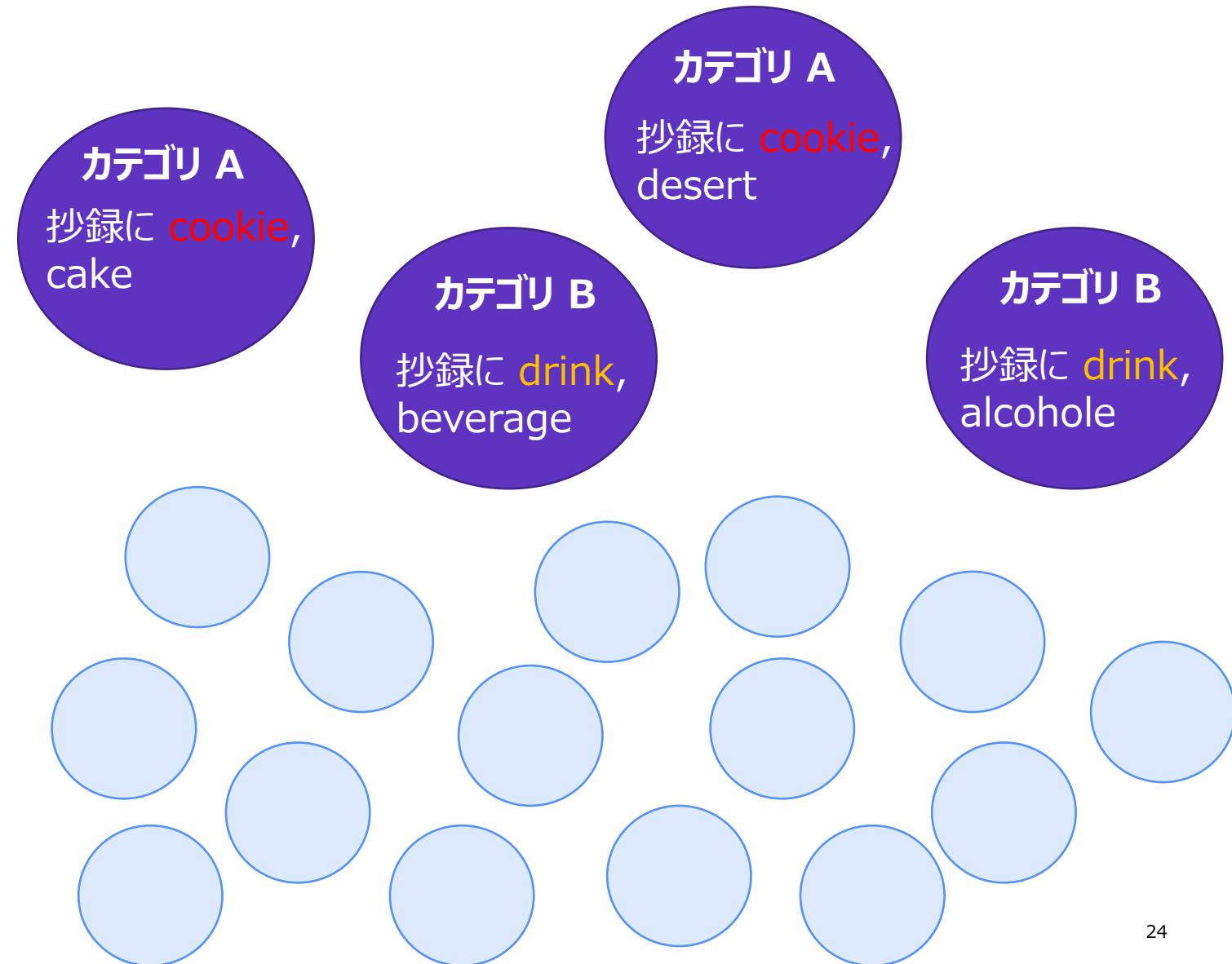
Derwent Data Analyzer:自動分類の仕組み



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

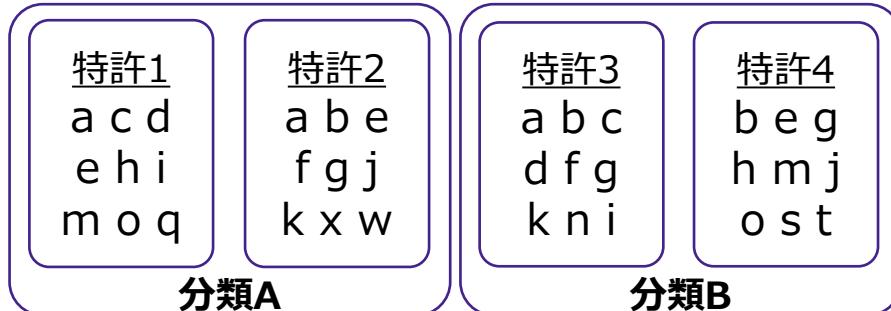


Derwent Data Analyzer:自動分類の仕組み

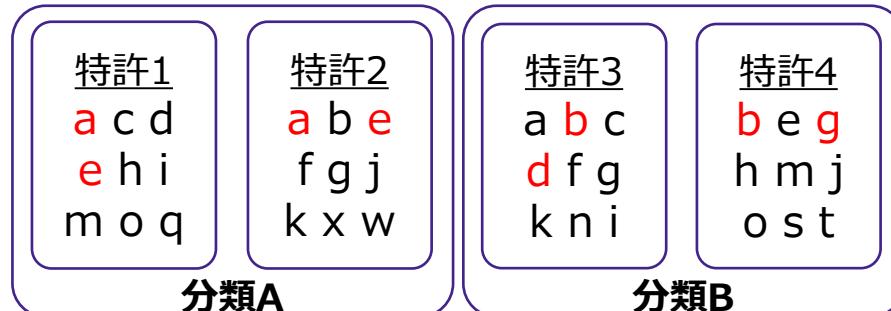
1. 機械学習プロセス

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



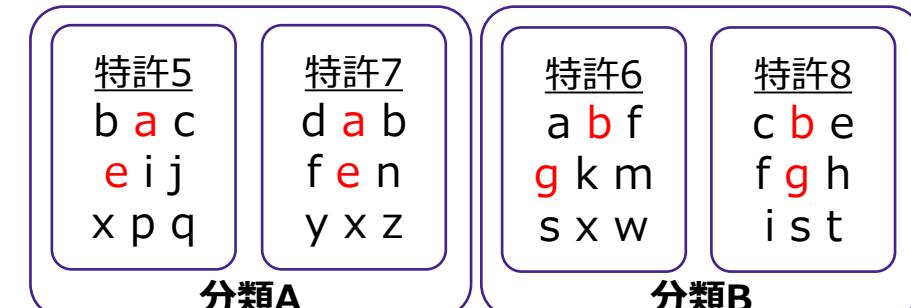
学習結果を活用

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



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

The screenshot shows the DDA software interface with the 'Analysis' tab selected. A purple circle labeled '1' points to the 'Record Classification Management' icon in the toolbar. The main window displays a 'Classification' dialog box. On the left, under 'Effect Category', a purple circle labeled '2' points to the 'Field Name Selection' section. On the right, under 'Category', a purple circle labeled '3' points to the 'Select the classification category' section. A callout bubble provides instructions for handling many categories via import. At the bottom of the dialog, a purple circle labeled '4' points to the 'Allow Multiple Selection' checkbox.

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)

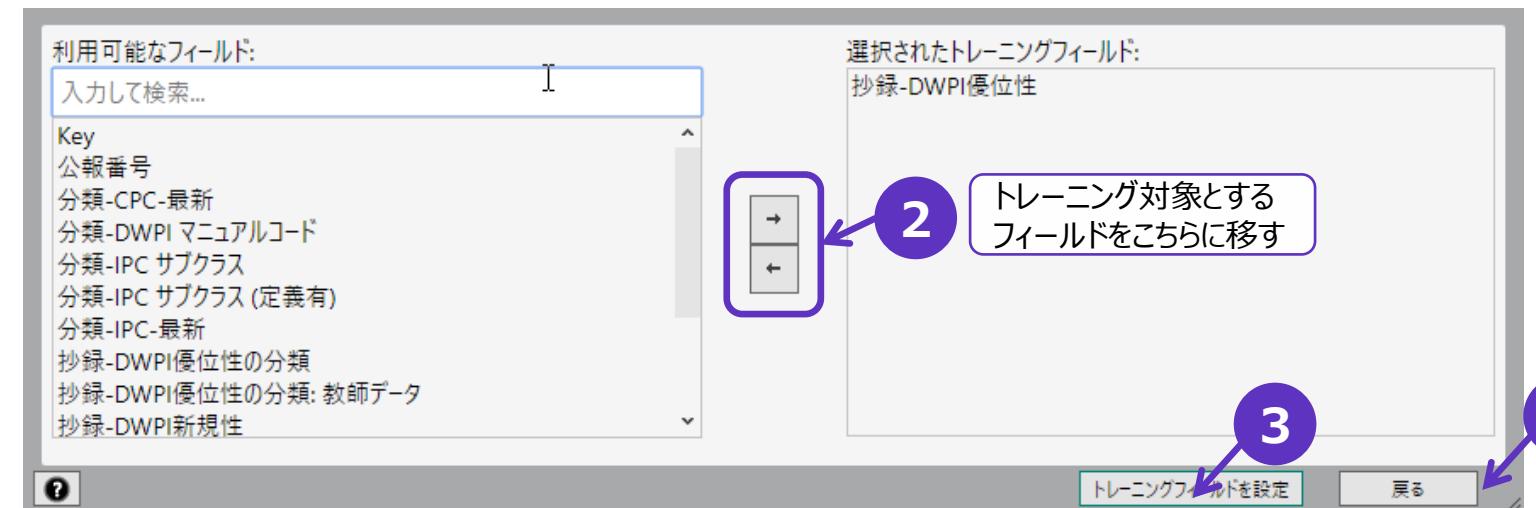
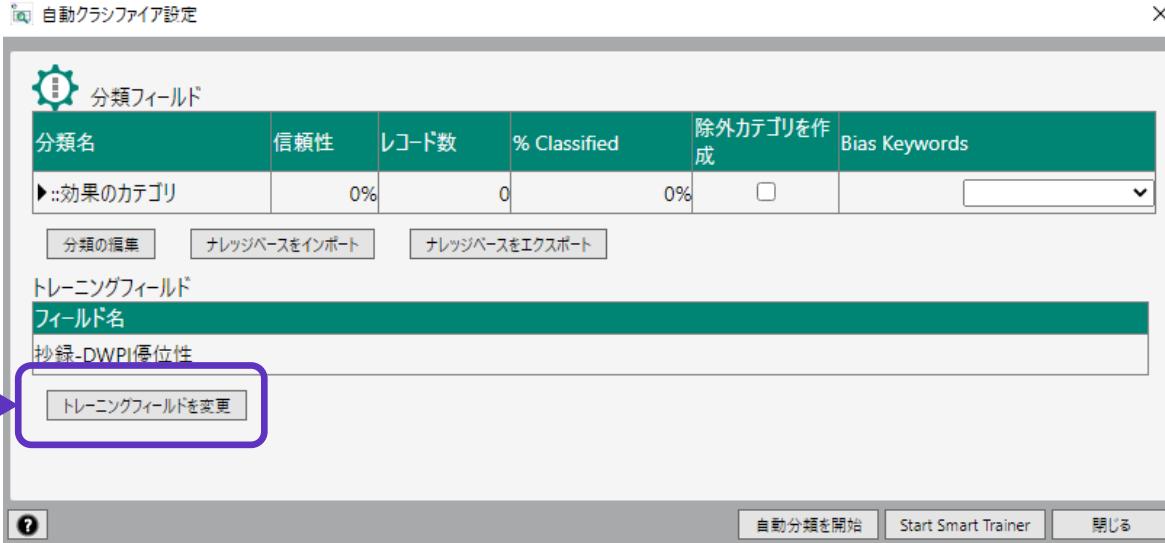
複数選択を許可
このチェックをONにすると、1つのレコードに対し、複数の分類を付与することが可能になります。複数の分類に当てはまる可能性がある場合は、チェックをONにしておきます。

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。

Clarivate™

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



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。

自動付与へ

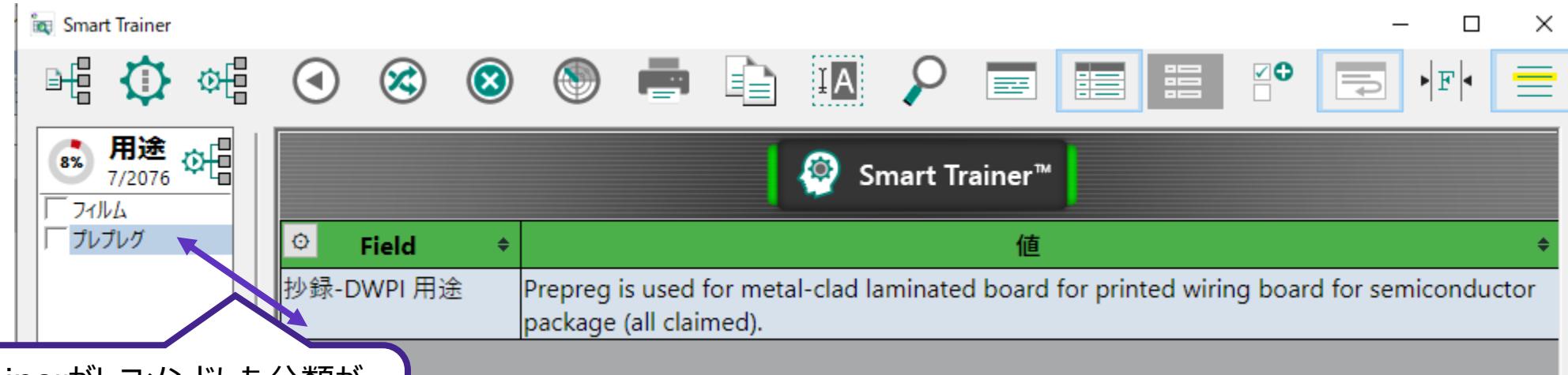
内容を読みながらカテゴリ付与

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。

Clarivate™

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。

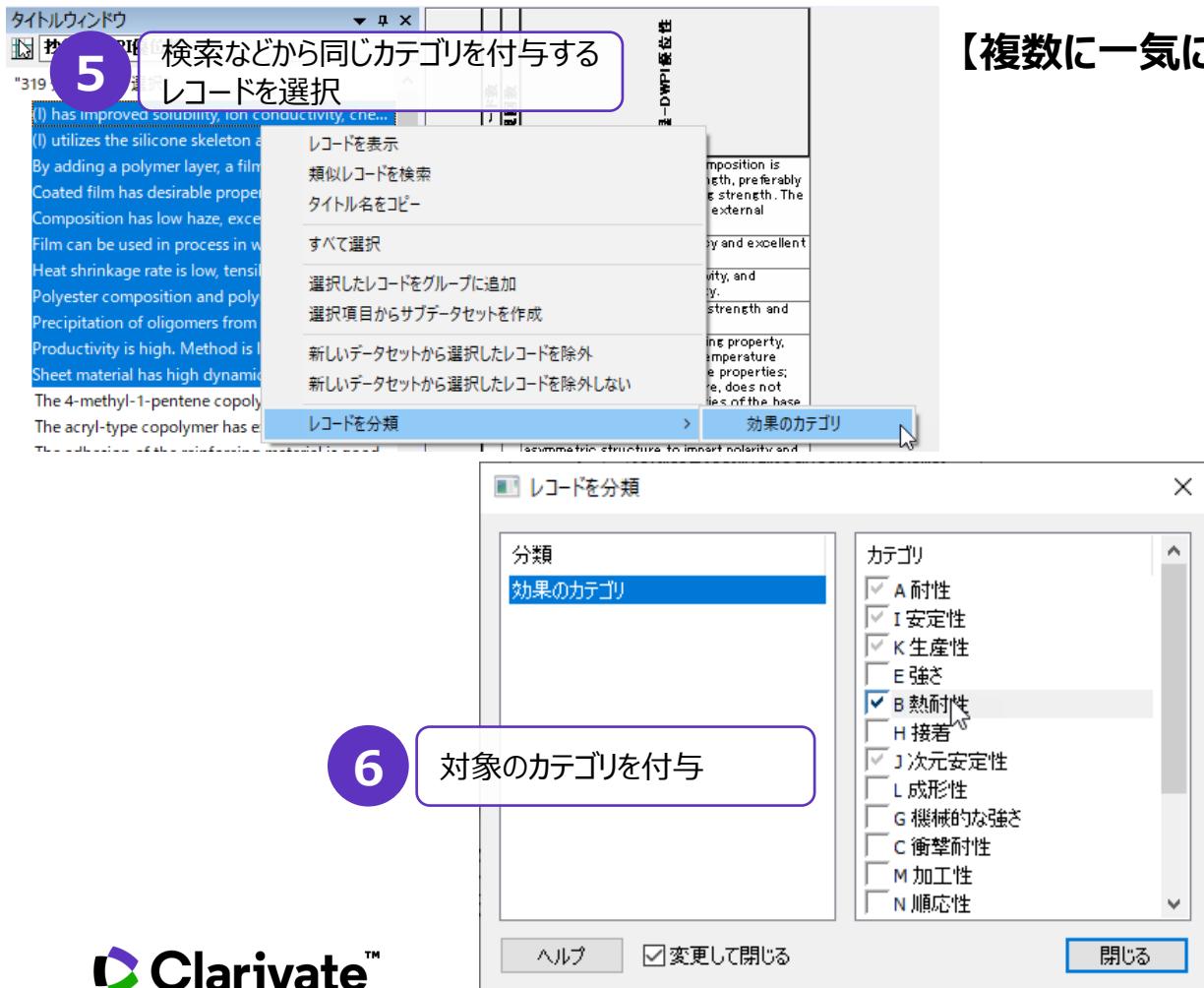


Smart trainerがレコメンドした分類が
ハイライトされます。
正しく学習していかどうか目安になります。

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

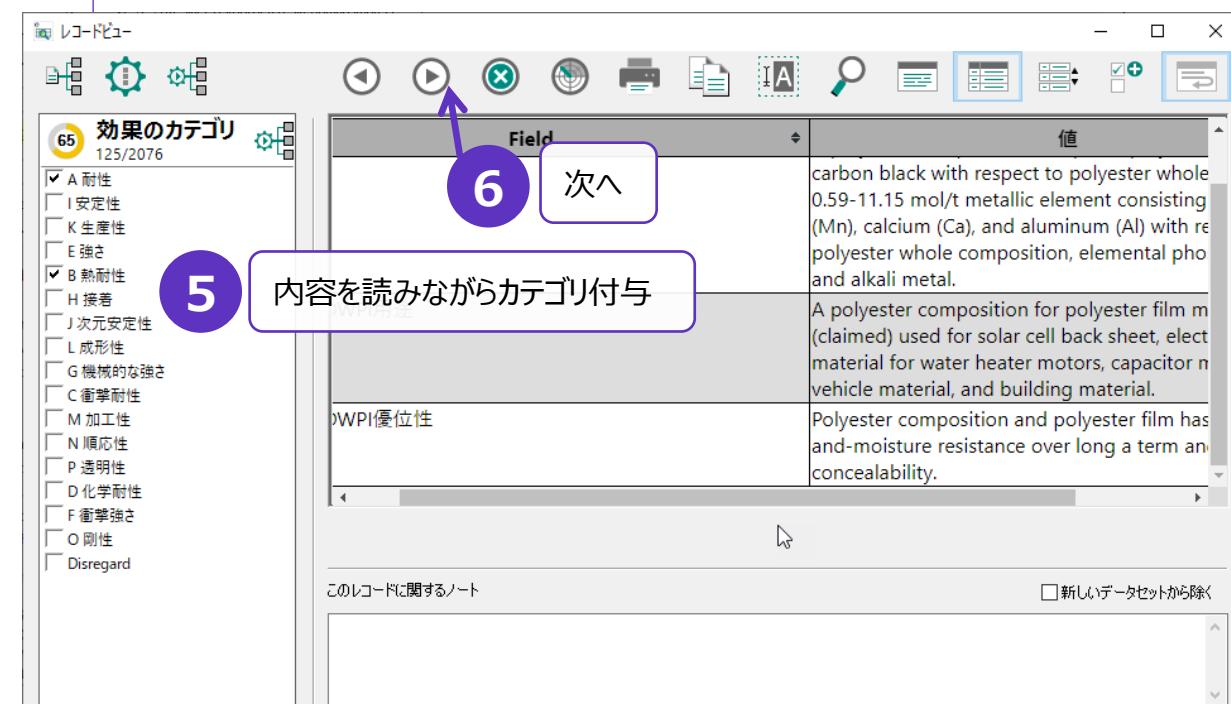


【複数に一気に付与】

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。

【レコード画面で1レコードずつ付与】



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

The screenshot shows the DDA software interface with several callouts and numbered steps:

- Callout 7:** Points to the "Smart Trainer" icon in the top menu bar.
- Callout 8:** Points to the classification results table in the "Automatic Classification" dialog box. The table lists categories like Disregard, 耐性, 安定性, 生産性, 強さ, 熱耐性, and 接着, along with their record counts and reliabilities.
- Callout 9:** Points to the "Start Classification" button in the "Automatic Classification" dialog box.
- Callout 10:** Points to the "Reliability Confirmation" box, which contains text about reliability values and training field changes.
- Callout 11:** Points to the "Precision Target" box, which discusses precision levels and category distribution.

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

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、自社でカテゴリ分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。

- ① スクリプトのインストールとメニューへの追加
- ② データ（教師データと未分類データ）のDDAへの取り込み
- ③ 分類フィールドの作成
- ④ トレーニングフィールドの設定
- ⑤ 教師データを読み込んで機械学習を行う
- ⑥ 未分類データへ自動で分類を付与する
- 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

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



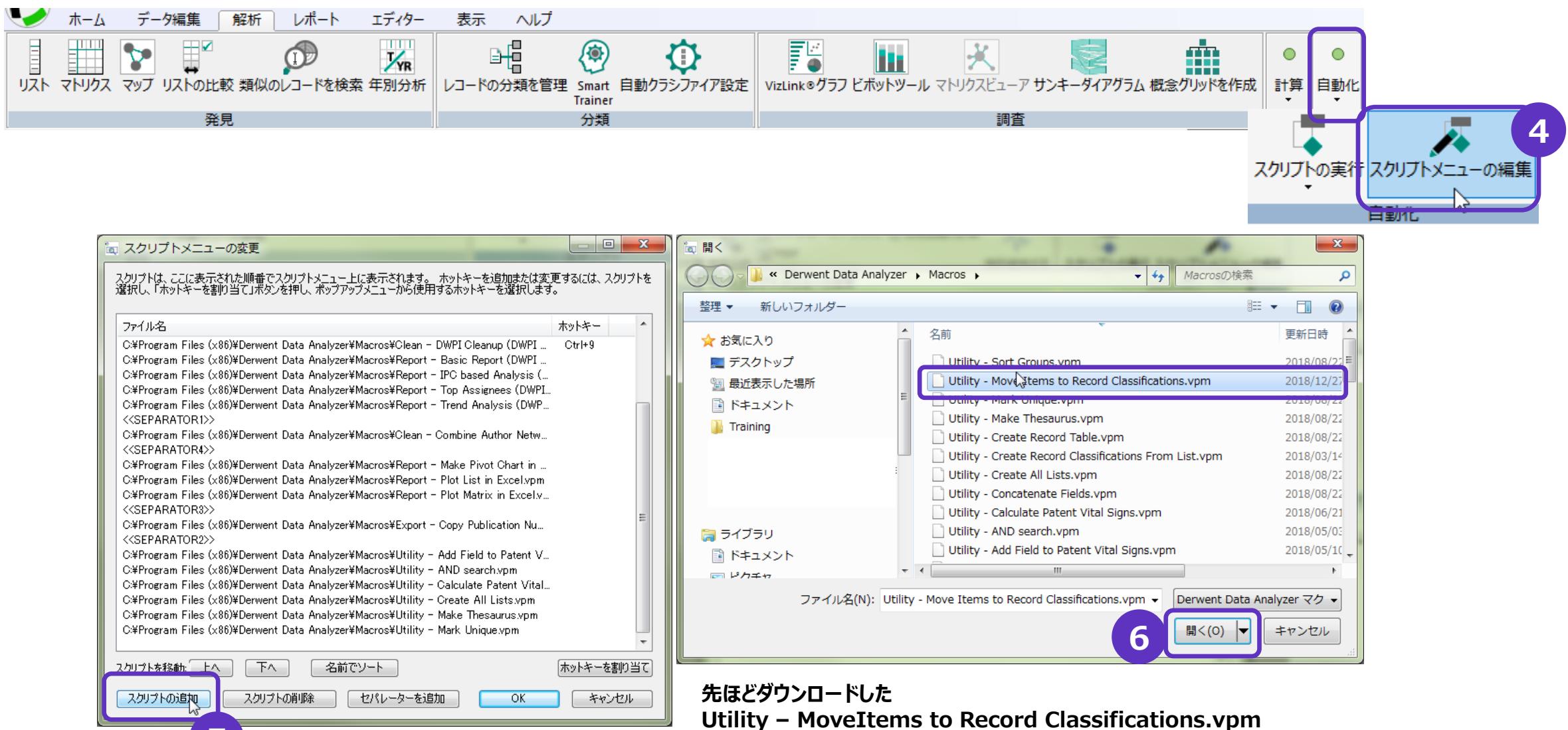
ダウンロードページが表示されます。

The screenshot shows the 'Downloads' section of the DDA interface. At the top, there are navigation links: 'Derwent Data Analyzer Home', 'System Requirements', 'Downloads' (which is highlighted with a purple oval and a number '2'), and 'Logout'. Below this, a breadcrumb trail shows 'Home > Downloads'. The main area is titled 'Downloads' and contains three tabs: 'Import Filters', 'Scripts' (which is highlighted with a purple oval and a number '2'), and 'Thesauri'. A purple circle labeled '3' points to a specific file entry in the list: 'Utility – Move Items to Record Classifications.zip'.

ダウンロードページが表示されます。

The screenshot shows the details for the 'Utility – Move Items to Record Classifications.zip' file. It includes columns for 'Name', 'Filename', and 'Last Update'. The 'Name' column shows 'Utility - Move Items to Record Classifications'. The 'Filename' column shows 'Utility - Move Items to Record Classifications.zip'. The 'Last Update' column shows '02/07/2019'. Below this table, there's an 'Install' button with an upward arrow icon, which is highlighted with a purple oval and a number '3'. A callout box also contains the text '3 Utility – Move Items to Record Classifications.zip を選択してインストール' (Select and install the script). At the bottom, there's a note: 'Requirements: New Record Classification field is already created before running the script.'

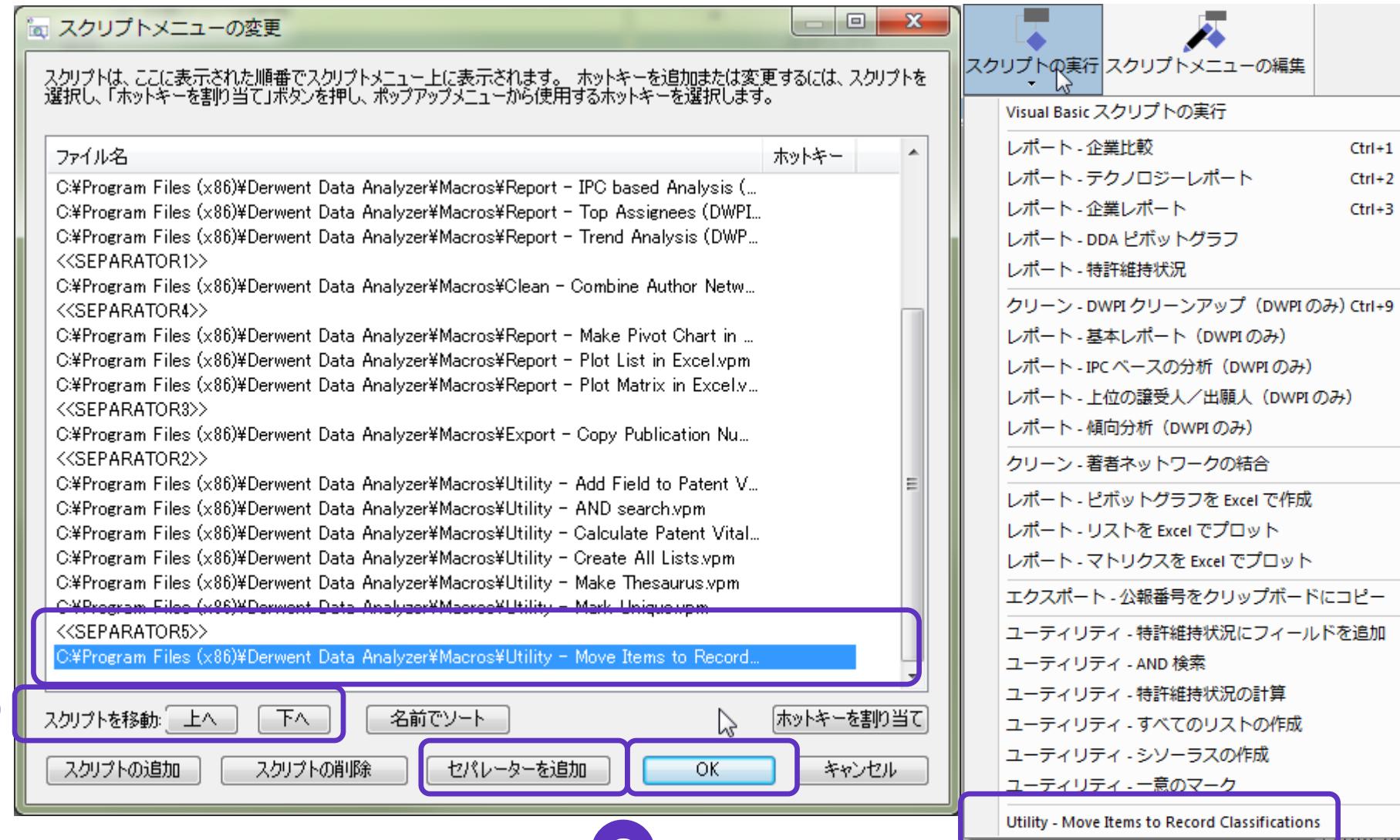
スクリプトをメニューに追加します



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。



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と一緒にした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と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-DWI新規	抄録-DWI新規性の分類	抄録-DWI優位性	抄録-DWI優位性の分類	抄録-DWI用途	抄録-DWI用途の分類
2 JP2013139523A	A polybutylene terephthalate film; H resin		Polybutylene terephthalate 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 carbon fiber; G 繊維; H 树脂		The carbon fiber has large A 耐性; E 強さ; 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 polymer A 耐性; B 熱耐性	Aromatic-polyamide porous A 自動車		
5 KR2013081453A	A white opaque polyester film; B ポリエチレン; F 組成物; H 树脂		The white opaque polyester E 強さ	White opaque polyester film D フィルム		
6 KR2013081124A	A heat-shrinkable polyimide film; A フィルム; B ポリエチレン; H 树脂		The heat-shrinkable polyimide A 耐性	Heat-shrinkable polyethylene film D フィルム		
7 KR2013080028A	Polyimide film (e), is A フィルム; I ポリイミド		The polyimide film: has I 安定性; K 生産性; J 次元安定性	The polyimide film is used D フィルム		
8 KR2013076391A	Manufacture of meta-aramid-based film		The meta-aramid-based film E 強さ	Manufacture of meta-aramid-based film C 製造; D フィルム		
9 KR2013076390A	Meta-aramid film composition		The film exhibits excellent E 強さ; G 機械的な強さ	The meta-aramid film is used D フィルム		
10 KR2013076044A	An adhesive protective film; A フィルム; F 組成物; H 树脂		The adhesive protection film A 耐性; H 接着	Adhesive protection film D フィルム		
11 KR2013074955A	Copper-clad laminate		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		The process: produces recycled 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 comprising a film; B ポリエチレン		The process enables efficient 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 base material		The carbon fiber base material E 強さ	Carbon fiber base material A 自動車; B プレプレグ; H 飛行機		
15 WO2013099608A1	A biaxially-oriented polyimide film; B ポリエチレン		The biaxially-oriented polyimide A 耐性; I 安定性; K 生産性; L 成形性	Biaxially-oriented polyimide film A 自動車; D フィルム; E モバイル		
16 WO2013099234A1	A polyphenylene sulfide film		The polyphenylene sulfide A 耐性; I 安定性; K 生産性; L 成形性	Polyphenylene sulfide film A 自動車		
17 WO2013099707A1	A sizing agent coated carbon fiber		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		The epoxy resin composition A 耐性; E 強さ; 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		The thermoplastic resin composition A 耐性; I 安定性; B 熱耐性	Thermoplastic resin composition is used for laser-marking A		
20 WO2013099607A1	A polyolefin microporous film		The polyolefin microporous film A 耐性; E 強さ	Polyolefin microporous film D フィルム		
21 WO2013099539A1	A polyolefin microporous film		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

The screenshot shows the Derive software interface with the 'Analysis' tab selected. A purple circle labeled '1' points to the 'Record Classification Management' icon in the toolbar. The main window displays a 'Classification' dialog box. On the left, under 'Effect Category', a purple circle labeled '2' points to the 'Field Name Selection' section. On the right, under 'Category', a purple circle labeled '3' points to the 'Select the classification category' section. A callout bubble provides instructions for handling many categories via import. At the bottom of the dialog, a purple circle labeled '4' points to the 'Allow Multiple Selection' checkbox.

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。

Clarivate™

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

The screenshot shows the 'Automatic Classifier Settings' dialog box and the 'Smart Trainer' ribbon bar.

1 The 'Training Field' section of the dialog box is highlighted with a purple border. It contains a table with one row: '抄録-DWPI優位性' (Summary-DWPI Dominance) under 'Field Name'. A button 'Training Field Change' is below it.

2 On the ribbon bar, the 'Classification' tab is selected. A callout bubble points to the 'Classification' icon in the top right corner of the ribbon bar, with the text 'ココからでも同じ画面を開けます。' (You can open the same screen from here.)

3 The 'Training Field Setting' button at the bottom right of the dialog box is highlighted with a purple border.

4 The 'Return' button at the bottom right of the dialog box is highlighted with a purple border.

This screenshot shows the 'Training Field Selection' dialog box.

1 The 'Available Fields' list on the left contains the following items:

- Key
- 公報番号
- 分類-CPC-最新
- 分類-DWPI マニュアルコード
- 分類-IPC サブクラス
- 分類-IPC サブクラス(定義有)
- 分類-IPC-最新
- 抄録-DWPI優位性の分類
- 抄録-DWPI優位性の分類: 教師データ
- 抄録-DWPI新規性

2 The 'Selected Training Fields' list on the right contains the item: '抄録-DWPI優位性' (Summary-DWPI Dominance). A button with arrows between the two lists is highlighted with a purple border.

3 The 'Training Field Setting' button at the bottom right of the dialog box is highlighted with a purple border.

4 The 'Return' button at the bottom right of the dialog box is highlighted with a purple border.

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

The screenshot shows the Clarivate software interface with the 'Classification' ribbon tab selected. The 'Classification' tab contains several icons: 'Record Classification Management', 'Smart Classifier Trainer', 'VizLink® Graph', 'Pivot Tool', 'Matrix View', 'Sankey Diagram', 'Concept Grid Creation', 'Calculation', and 'Automation'. The 'Automation' icon is highlighted with a purple circle.

1 Utility - Move Items to Record Classifications

2 学習させる分類のフィールドを選択してOK

3 自動分類させる先のフィールドを選択してOK

The 'Utility - Move Items to Record Classifications' window (step 1) lists various fields for selection. The '抄録-DWPI偽性の分類 教師データ' field is highlighted with a blue rectangle and a purple circle containing the number 2. The 'OK' button is also highlighted with a blue rectangle and a purple circle containing the number 2.

The 'Record Classification Selection' window (step 2) lists fields for assignment. The '改善点自動分類' field is highlighted with a blue rectangle and a purple circle containing the number 3. A callout bubble points to this field with the text '自動分類させる先のフィールドを選択してOK'. The 'OK' button is also highlighted with a blue rectangle and a purple circle containing the number 3.

The 'Record Classification Selection' window (step 3) lists fields for assignment. The '改善点自動分類' field is highlighted with a blue rectangle and a purple circle containing the number 3. A callout bubble points to this field with the text '自動分類させる先のフィールドを選択してOK'. The 'OK' button is also highlighted with a blue rectangle and a purple circle containing the number 3.

4 Automation Menu

The automation menu on the right side of the interface lists various report and utility options:

- 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
- レポート - 分類ツリーの参照
- 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

Clarivate™

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	出現回数	効果のカテゴリ	属性																	Disregard
			A 耐性	I 安定性	K 生産性	E 強さ	B 热耐性	H 接着	J 次元安定性	L 成形性	G 機械的強さ	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 開性		
1	1	1	14bf6a4f3v3r4ck4w59516ak4m26ru2n5o	✓			✓													
2	1	1	14hi5q716z1h5d3q5f6t41633n226r533o6k	✓																
3	1	1	1aty4n4e204d2360245e5s6py25z3s2b41g	✓																
4	1	1	1f48562i21211eb674q521vl5o26e1h3r1064	✓											✓	✓				
5	1	1	1v2k6k3yo37415n31a48124u281e716m5b6i20	✓	✓	✓														
6	1	1	1w20485j4w5e2o1n6h363w36l1f5x333i3455	✓				✓					✓							
7	1	1	235b2f11k3l3m593wk2e2w591a5t2l331n35	✓	✓	✓	✓													
8	1	1	25513flh3kn32541u5o445h6w1o1u3e5puk2y	✓																
9	1	1	2a3o673w1n172k115h1320534fb6s2ic6p324s	✓																
10	1	1	2fp4a5e7q3r4m6hb302f4f5f6g1y3h322a6w	✓																
11	1	1	2h131m2r1i5f4z256l6j2i2c6s1f5n3v2rx713s	✓	✓	✓														
12	1	1	2r4n4au1m1s52yl13f2is4h6m705i4p506x6u	✓																

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。
緑になつていれば、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"/>

分類の権重 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

?

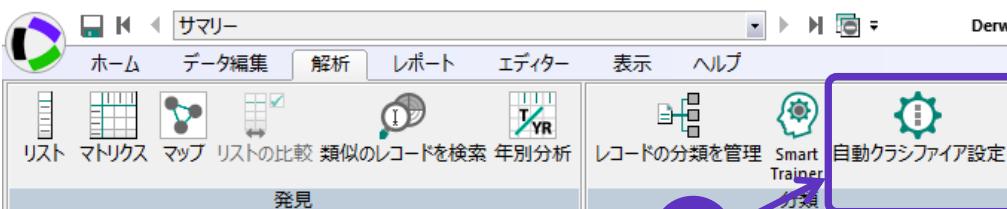
自動分類を開始 Smart Trainer を開始 閉じる

効果のカテゴリ	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 热耐性	33	55%
H 接着	20	58%

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

精度低めよりたくさんの中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 'Automatic Classification Settings' window. On the left, there's a table for 'Classification Fields' with one entry: '効果のカテゴリ' (Category) with 17 categories, 66% reliability, and 6% categorized. Below it are buttons for 'ナレッジベースをインポート' (Import Knowledge Base) and 'ナレッジベースをエクスポート' (Export Knowledge Base). On the right, a table lists categories with their record counts and reliabilities:

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

At the bottom, there are buttons: '自動分類を開始' (Start Automatic Classification), 'Smart Trainer を開始' (Start Smart Trainer), and '閉じる' (Close).

The screenshot shows the 'Automatic Record Classification' dialog box. It displays classification statistics: 65% reliability, 6.0% categorized, and the category '効果のカテゴリ' (Effect Category). There are buttons for 'すべて選択' (Select All), '選択なし' (No Selection), and 'アルファベット順にソート' (Sort Alphabetically). A progress bar shows '6 %' completion. At the bottom are buttons for '分類を開始' (Start Classification), 'ヘルプ' (Help), and '閉じる' (Close). Annotations explain the '精度目標' (Precision Target): a blue arrow points to the '精度目標' button with the text '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' (High - fewer records classified, fewer classification errors); another annotation with a red arrow points to the same button with the text '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' (Low - more records classified, more classification errors).

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 热耐性	H 接着	J 次元安定性	L 成形性	G 繊維的な強さ	M 加工性	N 順応性	P 透明性	D 化学耐性	F 繊維強さ	O 剛性	Unclassified
37	1	1	1	1 451 6n4h2d5w294u4c2z1d415u5q401k6c1m1j3l	<input type="checkbox"/>															
38	1	1	1	1 4544nk4x1l3ml24h6p3s5n1512c1p39k6e2c	<input type="checkbox"/>															
39	1	1	1	1 45b1w6m656sy5w2m5c2e5h2fp1d133b565mi	<input type="checkbox"/>															
40	1	1	1	1 46k712b1f524l6z1k2yl1q2q3k466n311v	<input type="checkbox"/>															
41	1	1	1	1 46m534f2f2g3z52473t4y2g4m115r5o3p3j2n	<input type="checkbox"/>															
42	1	1	1	1 4bf6a4t3v3r4ck4w59516ak4m26ru2n5o	<input type="checkbox"/>															
43	1	1	1	1 4h15q716z1h5d3q5f6t41633n226r533o6k	<input type="checkbox"/>															
44	1	1	1	1 4v4o5u3c6l113j5i5n6i4r6dl2w1j2y46182j	<input type="checkbox"/>															

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。

The screenshot shows the 'Record View' window with the title '27 改善点自動分類'. It lists several fields and their values, along with their corresponding classifications:

フィールド名	値
抄録-DWI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-DWI新規性の分類	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-DWI優位性	Thermoplastic resin composition is used for forming molded article

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													
		レコード数													
自動分類前		▲													
優先権主張年(最新)	レコード数	Show Values >= 1 and <= 32	共起レコード数	▼	▲	2008	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	72	C01:安全性	1	1	6	7	9	17	21	8	1	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)		1 5 58 68 68 12 16 98 9 4														
優先権主張年(最新)	レコード数	Show Values >= 1 and <= 54	共起レコード数	▼	▲	2008	2009	2011	2012	2013	2014	2015	2016	2017	2018	2019
1	142	C01:安全性				1	3	11	10	13	31	37	29	5	2	
2	182	C02:制御					9	25	20	38	54	33	2	1		
3	79	C03:油圧						15	11	9	15	23	5	1		
4	82	C04:表示	1	1	6	8	7	10	30	19						
5	133	C05:作業効率			1	18	14	10	31	42	15	2				
6	100	Unclassified	1	1	10	13	17	19	23	14	1	1				

リセット		優位性(インポート)															
		1 2 3 4 5 6 7 8 9 10 11															
自動分類後(2回目)		1 5 58 68 68 12 16 98 9 4															
優先権主張年(最新)	レコード数	Show Values >= 1 and <= 63	共起レコード数	▼	▲	2008	2009	2011	2012	2013	2014	2015	2016	2017	2018	2019	
1	149	C01:安全性				1	3	11	10	13	34	41	29	5	2		
2	211	C02:制御						13	28	25	42	63	37	2	1		
3	79	C03:油圧							15	11	9	15	23	5	1		
4	84	C04:表示	1	1	6	8	8	10	30	20							
5	152	C05:作業効率					1	20	18	13	36	43	18	3			
6	45	Unclassified	1	1	4	6	9	7	9	7							

606件のうち、561件を分類
(Unclassified除く)

※分類精度目標：高

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 回目自動分類後の結果をみてUnclassifiedに分類を付与

		レコード数	出発	属性	A01:シリンダ	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	106d6w2j2o4w1b6o439b4qd19363lx1v6mn									✓
2	1	1	161b1d445a2h6c1316m3m1t2p554h6e2f1z1u41									✓
3	1	1	16i4c3o2d433y516zm4b6o2v965yl52y636n									✓
4	1	1	181617435b1u1a3b3jf2j6t6c3t724k4aa4w5v									✓
5	1	1	1k6r2t3y3xa2d2e4z6h4s6c4h2s3r6u3t711t1s									✓
6	1	1	1p273m5t1x53921m2e6z2v313u3j68323t2y3t									✓
7	1	1	1vy4y5r686x1p6n4b476635h24132n3x523q6i									✓
8	1	1	275m1h1j3e1r40104l29303eyl4e44c59									✓
9	1	1	2b4w1715z5rf3al1l134m4x2n1y2194p									✓
10	1	1	2d2y4x6u221d5ft2a14h4n1m5l113g2g734s35									✓
11	1	1	2h66613j485914i2x42o2q3f3xl3128486n									✓

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	68	20	10	85	29	41

マトリクス（フレーズ×自動分類結果）												
新規性	レコード数	共起	レコード数	A01:シリンダ	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

※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用です。

(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。

The screenshot shows the software interface for record classification. On the right, a 'Record Classification' dialog box is open, showing a list of categories (A01:シリンダ, 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) with checkboxes. The 'A03:ブーム' checkbox is checked. Below the dialog is a 'Title Window' titled 'タイトルウィンドウ' with the identifier 'DWPI アクセション番号'. It displays a list of titles: "201650290M", "201723156Q", "201867735V", and "201896741T". A blue arrow points from the 'Title Window' towards the 'Record Classification' dialog. At the bottom of the interface, there is a message in Japanese: "Unclassifiedのものに他の分類を付与します" (Assign other categories to Unclassified items).

(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。

The screenshot shows a software interface for patent record review. On the left, there is a sidebar titled "優位性" (Priority) with a count of 78 and 606/606 total. It contains several classification categories with checkboxes: C01:安全性, C02:制御, C03:油圧, C04:表示, and C05:作業効率 (which is checked). Below this is an "Unclassified" option. The main area displays a table with columns "Field" and "値" (Value). The table includes fields such as 公報番号 (Derwent Innovation リンク), DWPI アクセッショング番号, ベーシック特許, タイトル (ベスト), 引用(Forward) 特許数-DPCI, DWPI ファミリー国数, DWPI ファミリーメンバー数, 独立請求項 (数), 謙受人/出願人 (ベスト), 発明者 (クリーン), DWPI クラス (定義有), and 分類-IPC-最新 (定義有). The "Title (Best)" field contains a detailed description of the invention. The "DWPI Class (Defined)" field lists Q38: Mechanical -> Conveying, Packaging, Storing -> Hoisting; Lifting; Hauling; Trucks. The "Classification-IPC-Latest (Defined)" field lists B66C 23/40: Performing Operations; Transporting -> Hoisting; Lifting -> 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 -> mounted on arms, 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。

The screenshot shows a software interface for patent record review. At the top, there is a ribbon menu with tabs: ホーム, データ編集, 解析, レポート, エディター, 表示, ヘルプ. The "解析" tab is highlighted. Below the ribbon, there are several icons: リスト (List), マトリクス (Matrix), マップ (Map), リストの比較 (List Comparison), 類似のコードを検索 (Search Similar Code), 年別分析 (Year-by-Year Analysis), and 発見 (Discovery). The "Smart Trainer" icon is highlighted with a blue box. To its right, there is a sidebar titled "新規性" (Novelty) with a count of 75 and 606/606 total. It contains several classification categories with checkboxes: A01:シリンダー, A04:ジブ, A02:アウトガ, A03:ブーム, A05:ディスプレイ, A06:ポンプ, A07:アーム (which is checked), and Unclassified. The main area displays a table with columns "Field" and "値" (Value). The table includes a field labeled "抄録-DWPI 新規性: NLP/フレーズ" which contains a long string of keywords separated by double pipes: 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

The screenshot shows the Clarivate software interface with the following steps highlighted:

- 1** In the top ribbon, click the 'DWPI' button, then click 'Group from Group'.
- 2** In the 'Group Name from Group' dialog, click the 'Group Name' tab, then select the 'Priority' field.
- 3** In the 'Group Name from Group' dialog, check the boxes for 'C01: 安全性', 'C02: 制御', 'C03: 油圧', 'C04: 表示', and 'C05: 作業効率'.

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	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

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

The screenshot shows the Derwent Classification tool interface. The top menu bar includes 'ホーム' (Home), 'データ編集' (Data Edit), '解析' (Analysis), 'レポート' (Report), 'エディター' (Editor), '表示' (View), and 'ヘルプ' (Help). The toolbar contains icons for 'リスト' (List), 'マトリクス' (Matrix), 'マップ' (Map), 'リストの比較' (List Comparison), '類似のレコードを検索' (Search Similar Records), '年別分析' (Year-by-Year Analysis), 'レコードの分類を管理' (Manage Record Classification), 'Smart Trainer' (Smart Trainer), and '自動クラシファイア設定' (Automatic Classifier Setting). A purple circle labeled '1' points to the '自動クラシファイア設定' icon.

A large central window titled '自動クラシファイア設定' displays a '分類フィールド' (Classification Field) table:

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"/>

Below the table are buttons: '分類の編集' (Edit Classification), 'ナレッジベースをインポート' (Import Knowledge Base), and 'ナレッジベースをエクスポート' (Export Knowledge Base). A purple circle labeled '2' points to the 'ナレッジベースをエクスポート' button.

A modal dialog titled 'ナレッジベースをエクスポート' (Export Knowledge Base) is open:

エクスポートする分類を選択: **優位性**

エクスポートするフィールド: **抄録-DWPI 優位性: NLP/フレーズ**

Ok キャンセル

A purple circle labeled '3' points to the '優位性' field in the export selection dropdown.

To the right of the main window, a detailed table titled '優位性 カテゴリ' (Priority Category) shows the distribution of records:

カテゴリ	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%

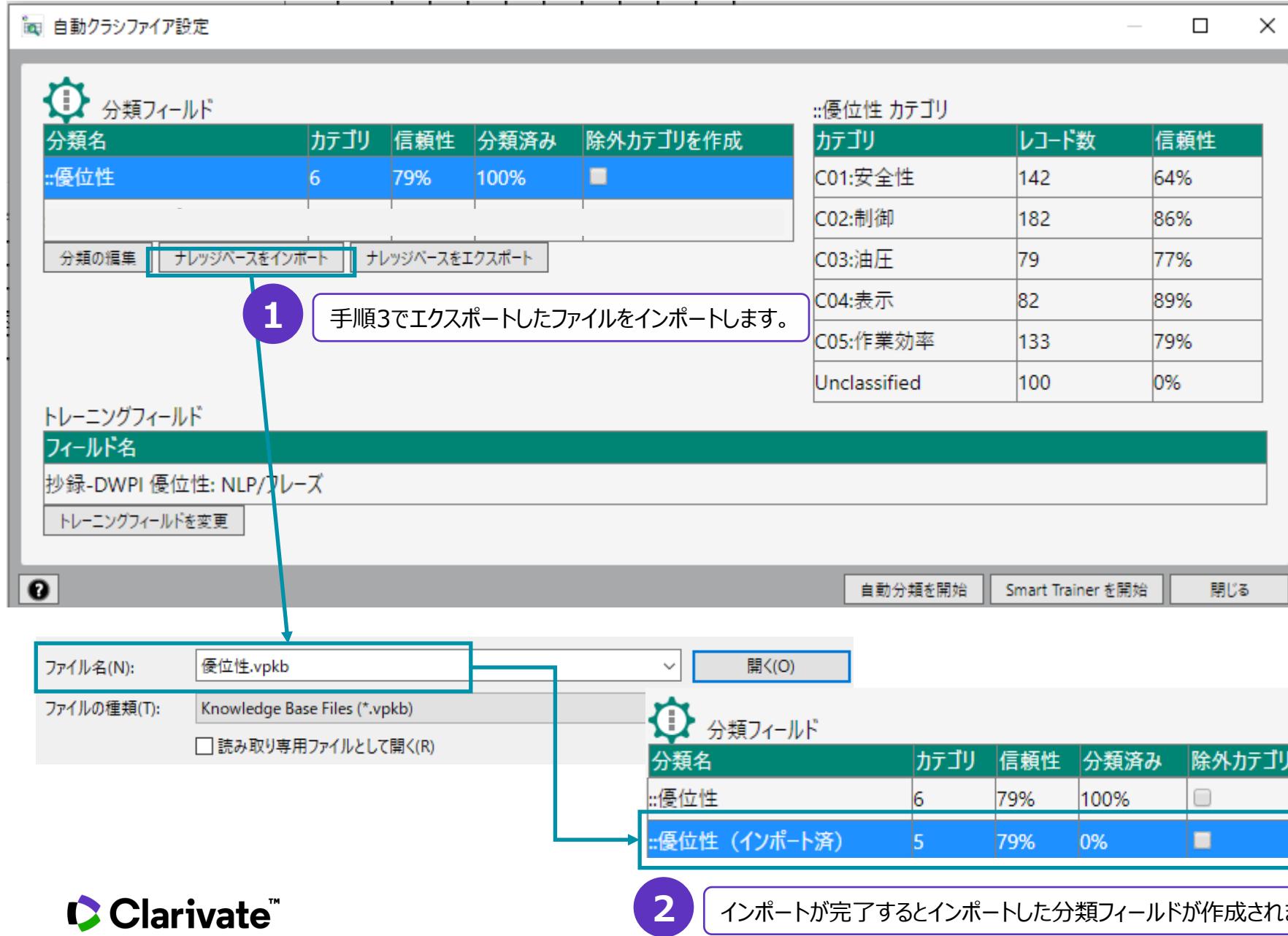
A purple circle labeled '4' points to the text: 'エクスポートしたデータを拡張子vpkbで保存します。' (The exported data is saved with the extension vpkb).

At the bottom left, there is a note: 'ナレッジベースをエクスポートをクリックし、エクスポートするフィールドを選択します。' (Click 'ナレッジベースをエクスポート' and select the export field).

At the bottom right, there is a note: '気になる分類フィールドをクリックすると各カテゴリへの分類比率を確認できます。' (Click on a classification field to check the classification ratio for each category).

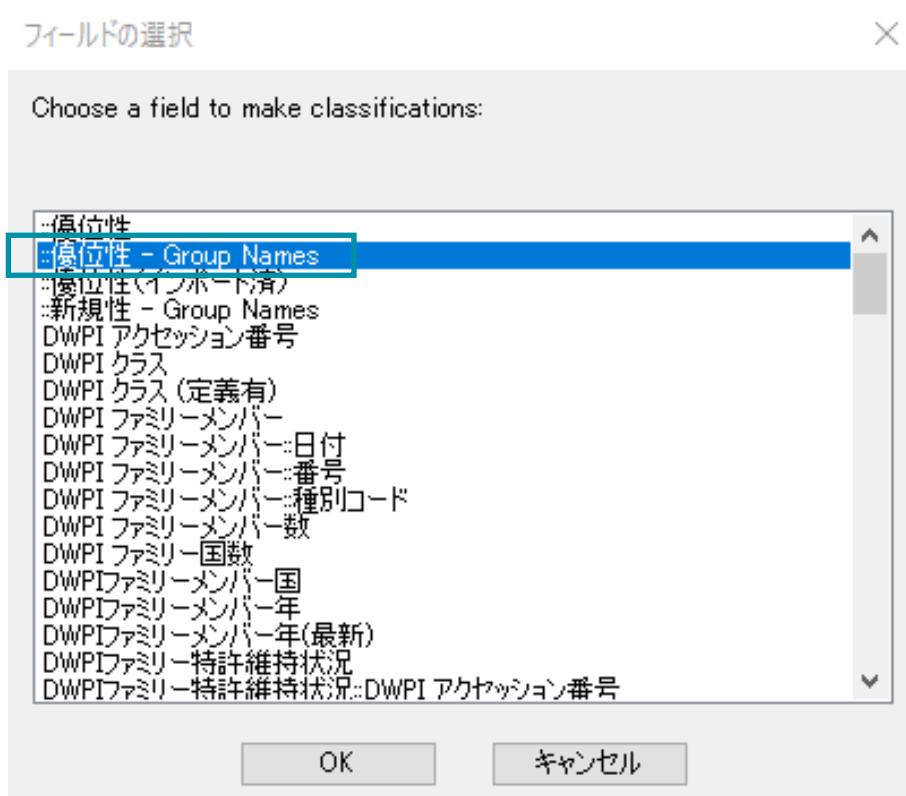
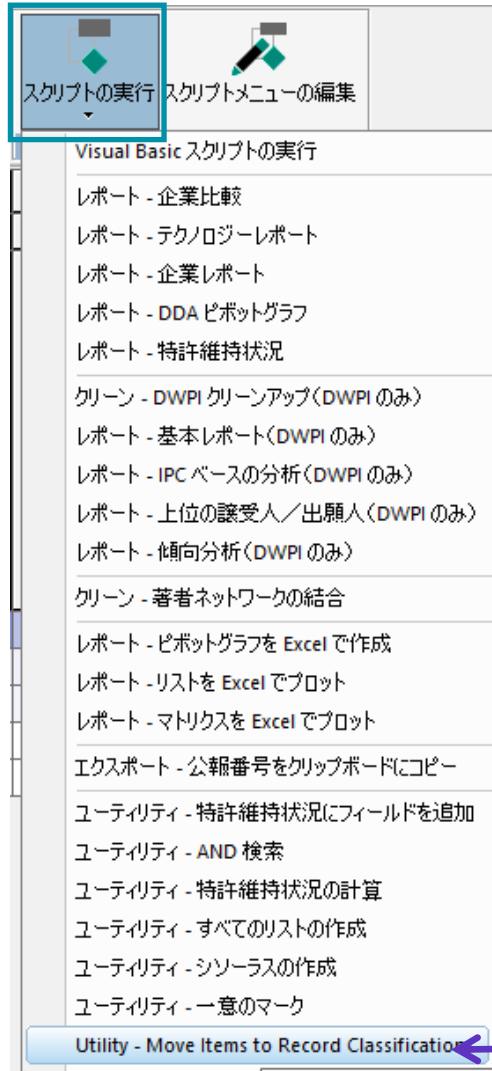
The bottom navigation bar includes buttons for '自動分類を開始' (Start Automatic Classification), 'Smart Trainer を開始' (Start Smart Trainer), and '閉じる' (Close).

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



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。

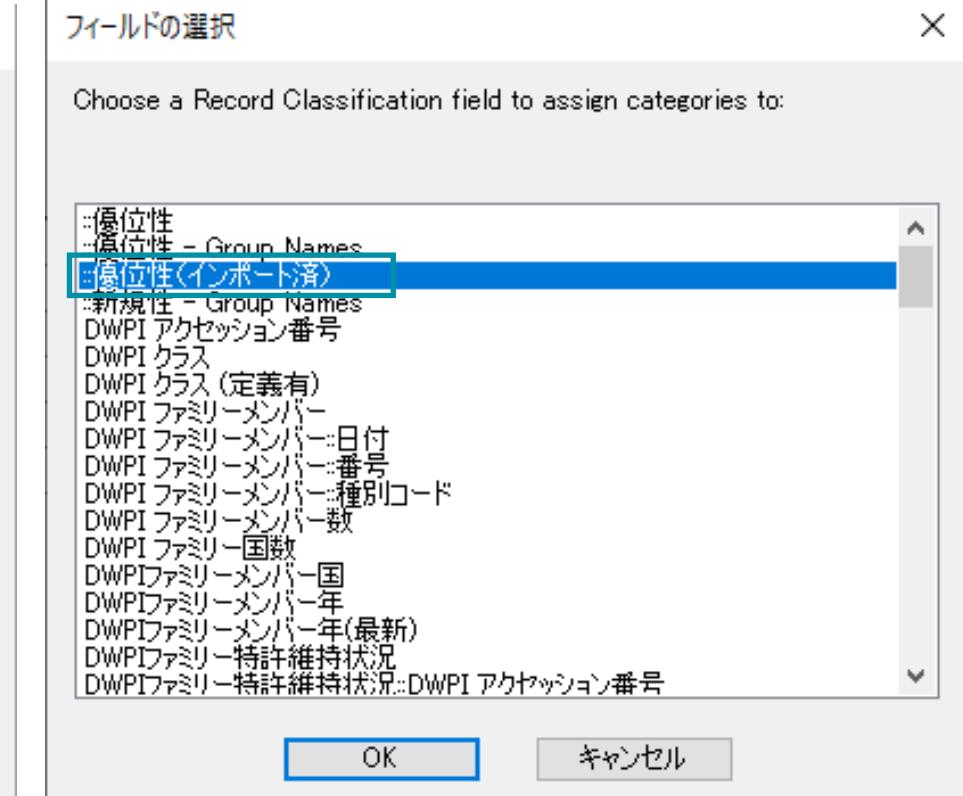
5 手順 4 でインポートして作成した分類カテゴリで再分類を行う



2

手順 2 で作成したフィールドを選択します。

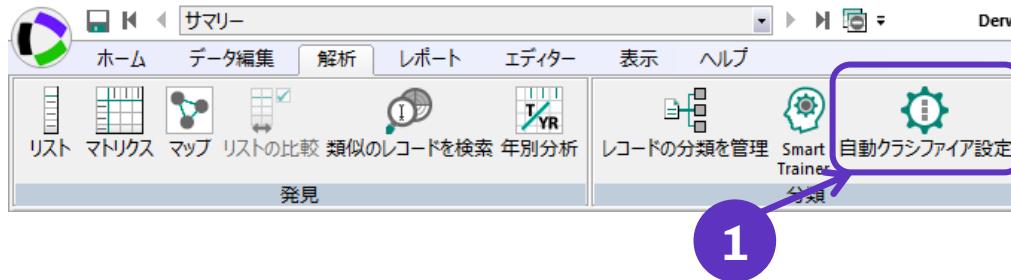
1



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
自動クラシファイア設定で確認することができます。

This dialog box displays classification results for a category named '優位性 (インポート済)'. It includes a table of categories and their counts, and a table of training fields. A purple circle labeled '2' points to the second table.

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)	6	76%	100%	<input checked="" 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	149	75%
C02:制御	211	77%
C03:油圧	79	77%
C04:表示	84	78%
C05:作業効率	152	75%
Unclassified	45	0%

トレーニングフィールド		
フィールド名	抄録-DWPI 優位性: NLP/フレーズ	
トレーニングフィールドを変更	<input type="button" value="トレーニングフィールドを変更"/>	

底部には「自動分類を開始」、「Smart Trainer を開始」、「閉じる」ボタンがあります。

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。

This dialog box allows selecting classification fields and setting accuracy goals. A purple circle labeled '3' points to the 'Start Classification' button.

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 type="checkbox"/> 78%	100.0%	優位性
<input checked="" type="checkbox"/> 76%	100.0%	優位性(インポート済)

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

すべて選択 選択なし アルファベット順ソート

自動クラシファイアの進捗
100 %

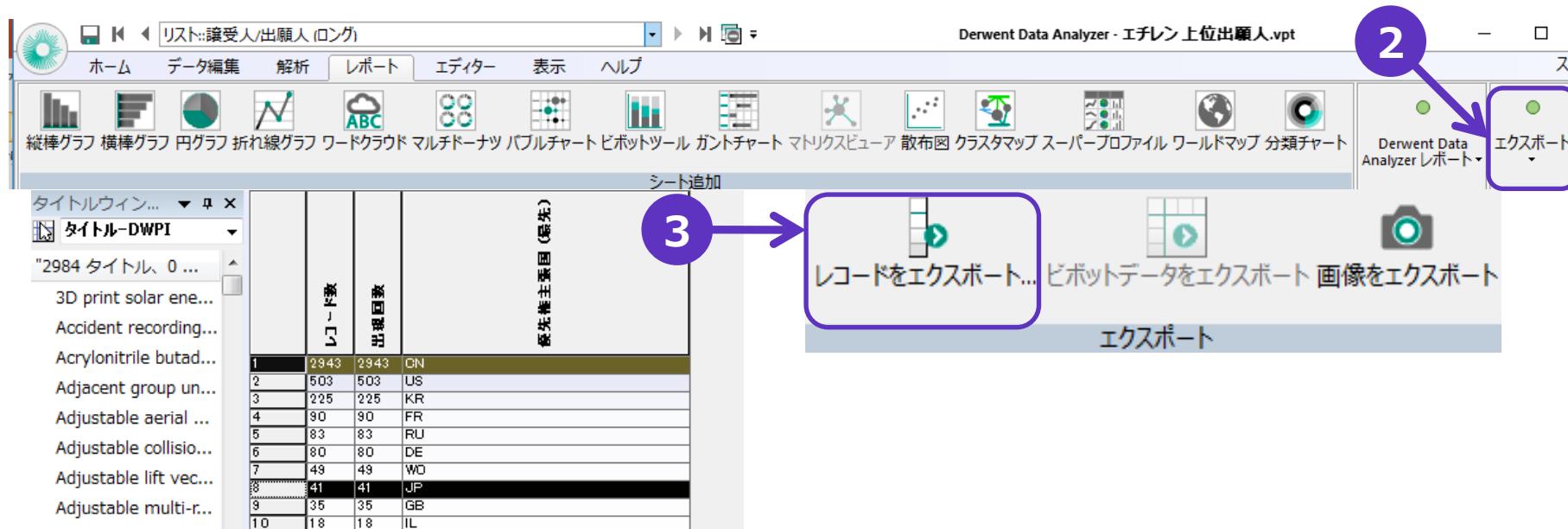
分類を開始 ヘルプ 閉じる

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上で様々なデータ整理・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。

2

エクスポート

3

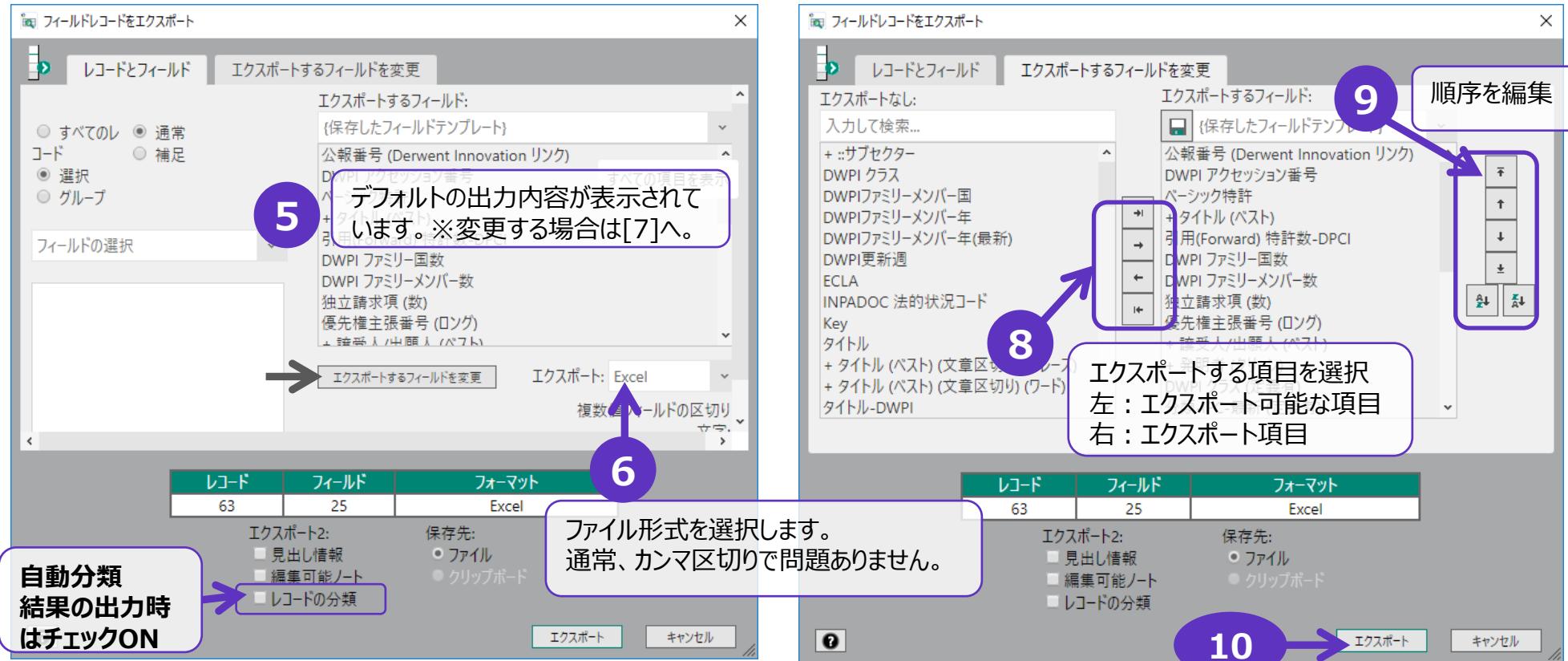
レコードをエクスポート... ピボットデータをエクスポート 画像をエクスポート

エクスポート

Excelへのデータ出力

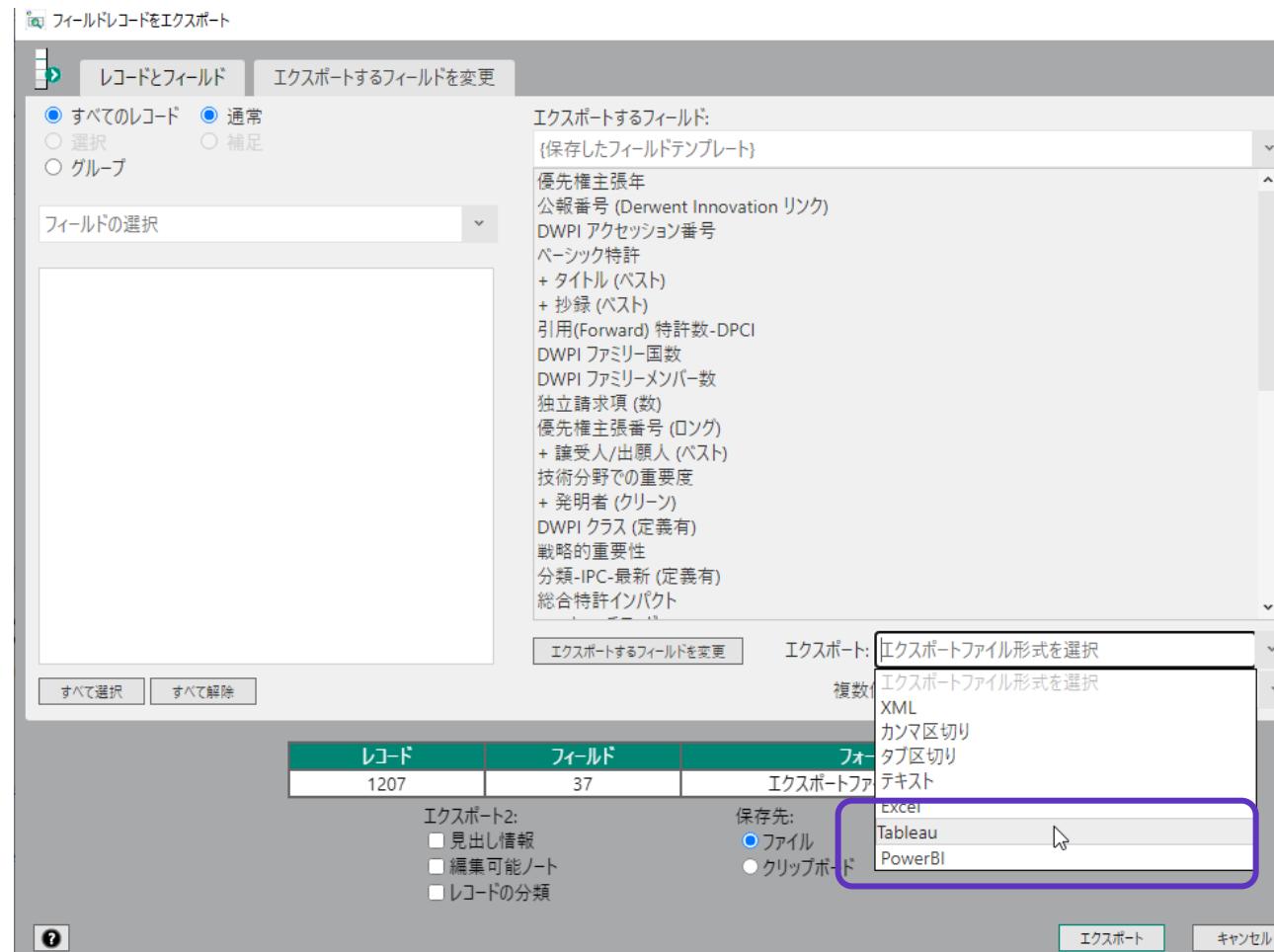
出力したデータは、再度DDAで分析作業を行うために、DDAに取り込むことが可能です。CSVファイルをExcel形式で保存します。※自動分類（レコードの分類）を含んで出力する場合は、CSVファイルのみ対応しています。

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。



エクスポートフォーマットTableauとPowerBIを追加

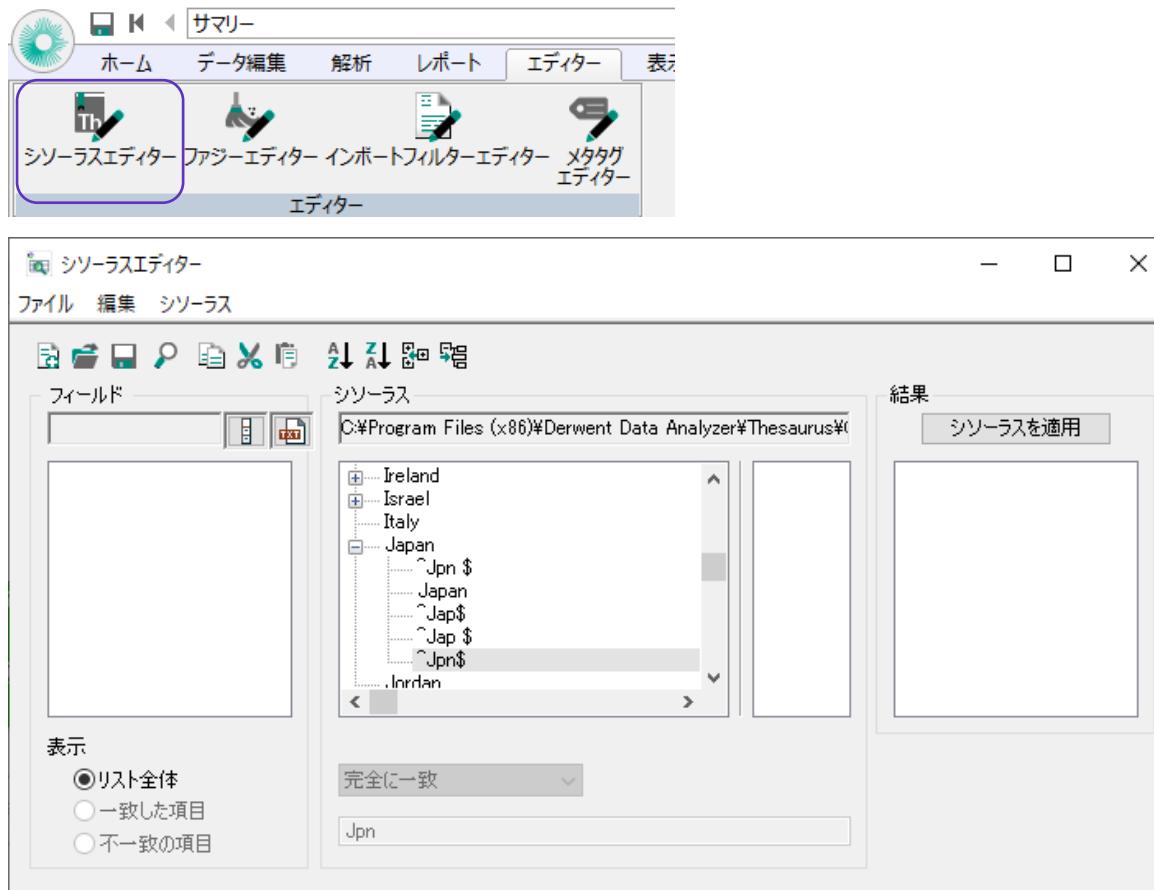
- ◆ Tableau や PowerBIへデータをインポートするためのエクスポートフォーマットを追加



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

A screenshot of a text editor window titled 'Country.the - メモ帳'. The content of the file is:

```
#Country/Region thesaurus
#Description: Country/Region thesaurus
#Description-ja-JP: 国/地域シソーラス
#Description-zh-CN: 国家/地区叙词表
**Afghanistan
100 1 ^Afghan
**Algeria
**Antigua
100 1 ^Barbuda
**Argentina
100 1 ^Argent
100 1 ^Arg$
**Australia
100 1 ^Aust $
100 1 ^Aust$
**Austria
100 1 ^Austr$
100 1 ^Austr $
**Azerbaijan
```

The text editor window includes standard controls like minimize, maximize, and close buttons, and status bars at the bottom indicating '1行、1列' (1 line, 1 column), '100%', 'Windows (CRLF)', and 'UTF-8 (BOM付き)'.

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 $atomic force microscope probe$b↵  
**Auto-focus↵  
100 1 $auto focus$b↵  
100 1 $automatic focus$b↵  
100 1 $automatic focusing$b↵  
**lithium-ion-cell↵  
100 1 $lithium ion battery$b↵  
100 1 $lithium ion batteries$b↵  
100 1 $lithium-ion battery$b↵  
100 1 $lithium-ion batteries$b↵  
100 1 $lithium-ion secondary batteries$b↵  
100 1 $lithium ion secondary batteries$b↵  
100 1 $lithium-ion secondary battery$b↵  
100 1 $lithium ion secondary battery$b↵  
100 1 $lithium-ion cell$b↵  
100 1 $lithium ion cell$b↵  
100 1 $lithium ion secondary cell$b↵  
100 1 $lithium-ion secondary cell$b↵  
**pedestrian↵  
100 1 $walker$b↵  
**vehicle↵  
100 1 $vehicles$b↵  
100 1 $car$b↵  
100 1 $cars$b↵  
100 1 $automobile$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 mic
9	2	2	Auto-focus microscope for inspecting specimen has decision unit ↴
10	2	2	Automatic focus adjustment in microscope computes focussing rai
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

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

