Clarivate™

Derwent Data Analyzer マニュアル③

~ファイルとデータの操作・機械学習編~

Clarivate 2026年

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

Clarivate™

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

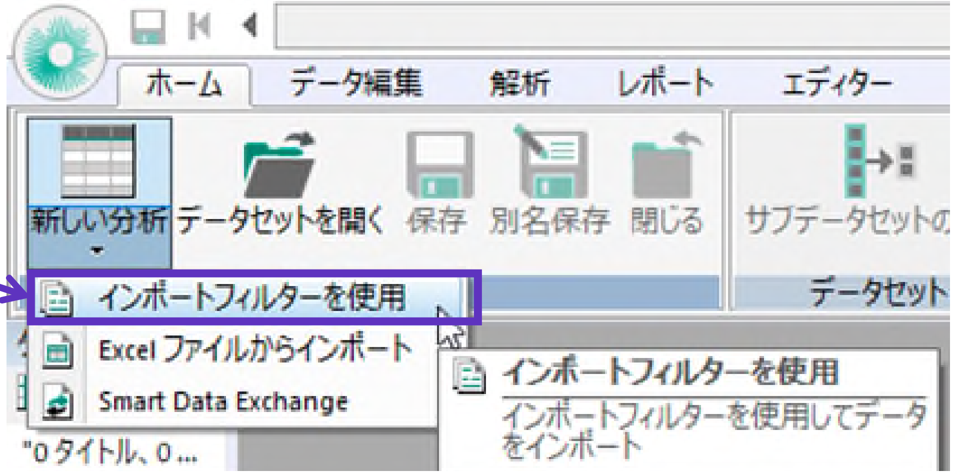
Web of Science、INSPEC等、同じ方法で行うことができます。
エクスポートの方法についての詳細は、Web of Scienceのマニュアルを参照してください。

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

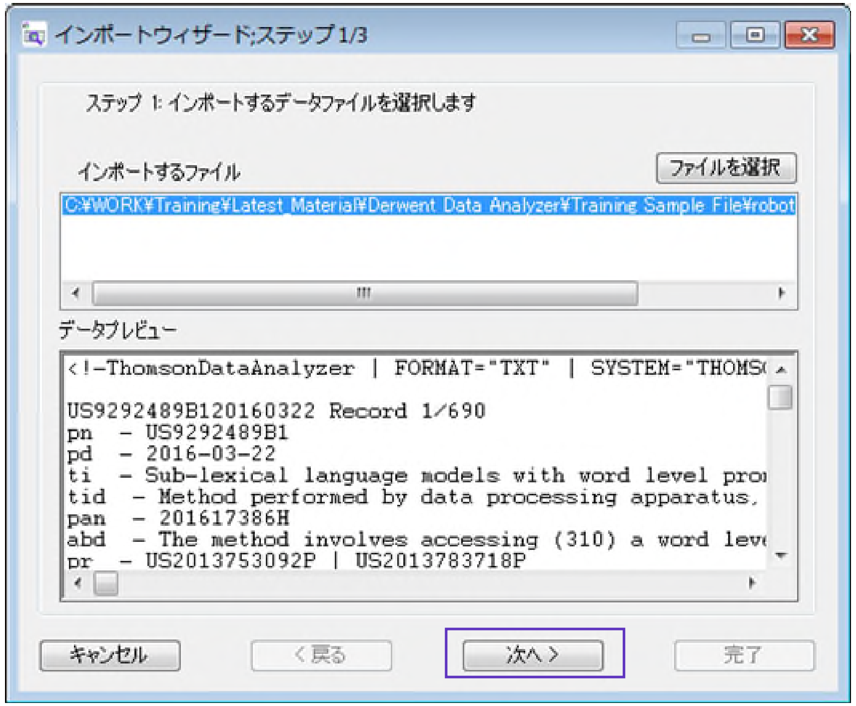
2 ファイルを適切な場所に保存

Web of Scienceの学術文献データを取り込む

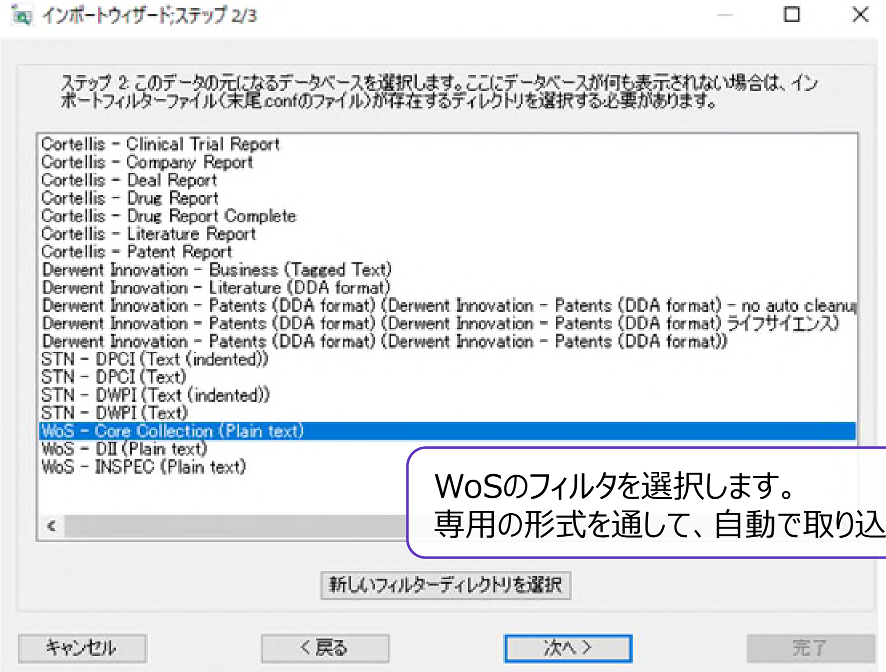
1



2



3



WoSのフィルタを選択します。
専用の形式を通して、自動で取り込みます。

Clarivate™

Excel形式のデータをDDAに取り込む

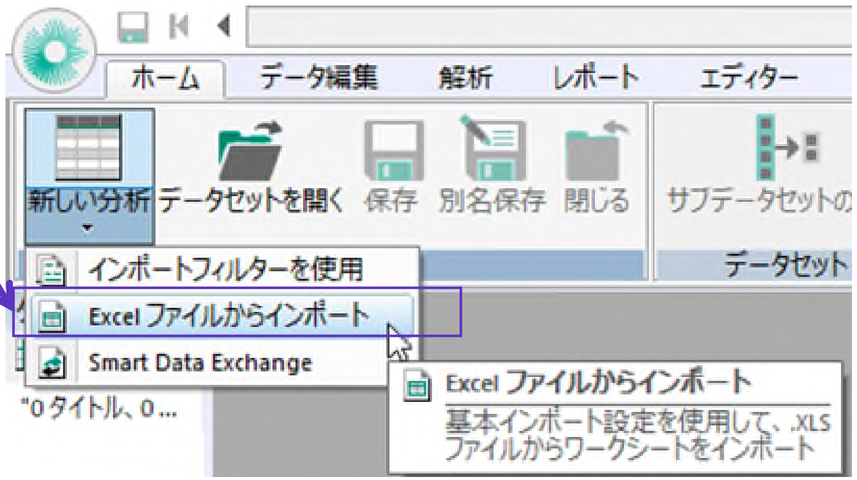
下の例のように、最初の行に各列名があり、2行目以降に各レコードの該当のフィールド情報が入った形式ファイルであれば利用可能です。
様々なデータベースでダウンロードした情報をDDAにインポートすることが可能です。

【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_A α 5 receptor - useful for the treatment of diminished					
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.	Alzheimers 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 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 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.	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;	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 PPARα agonists - useful for the treatment of lipid disorders and for reduction of the risk of development of atherosclerosis.	Alzheimers disease; Atherosclerosis; Crohns disease; Dermatological disease; Hypercholesterolemia; Hyperglycemia; Hyperlipidemia; Irritable bowel syndrome; Lipid metabolism disorder; Non-insulin dependent diabetes; Obesity; Restenosis;	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 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.	Acquired Immune Deficiency Syndrome; 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

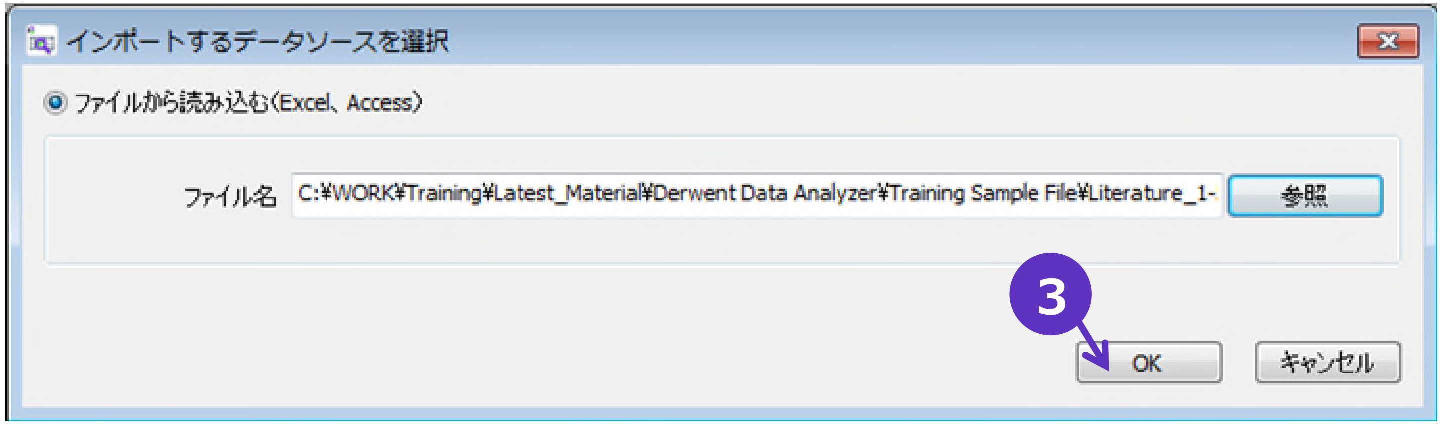
Excel形式のデータをDDAに取り込む：Excelデータをインポート

1



2

インポートするExcelファイルを選択します。



3

Clarivate™

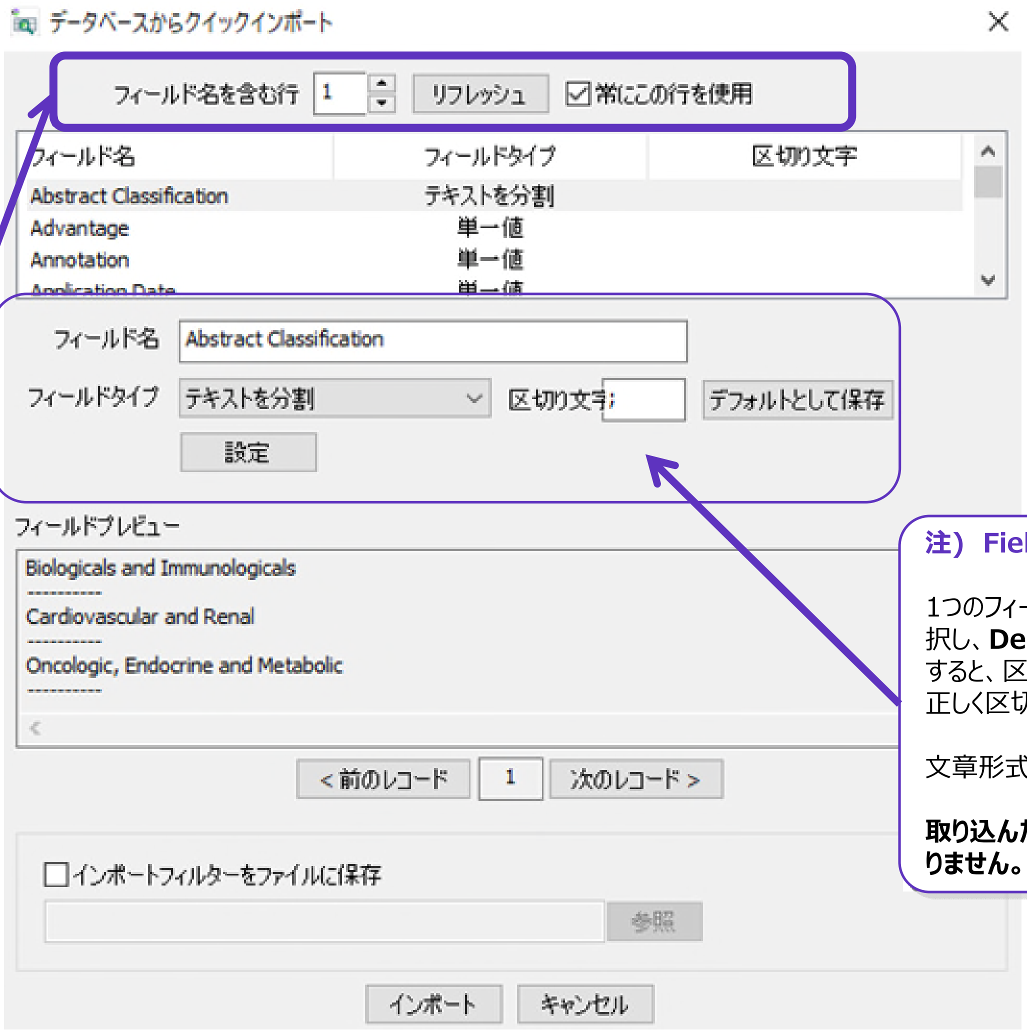
Excel形式のデータをDDAに取り込む：Excelデータをインポート

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

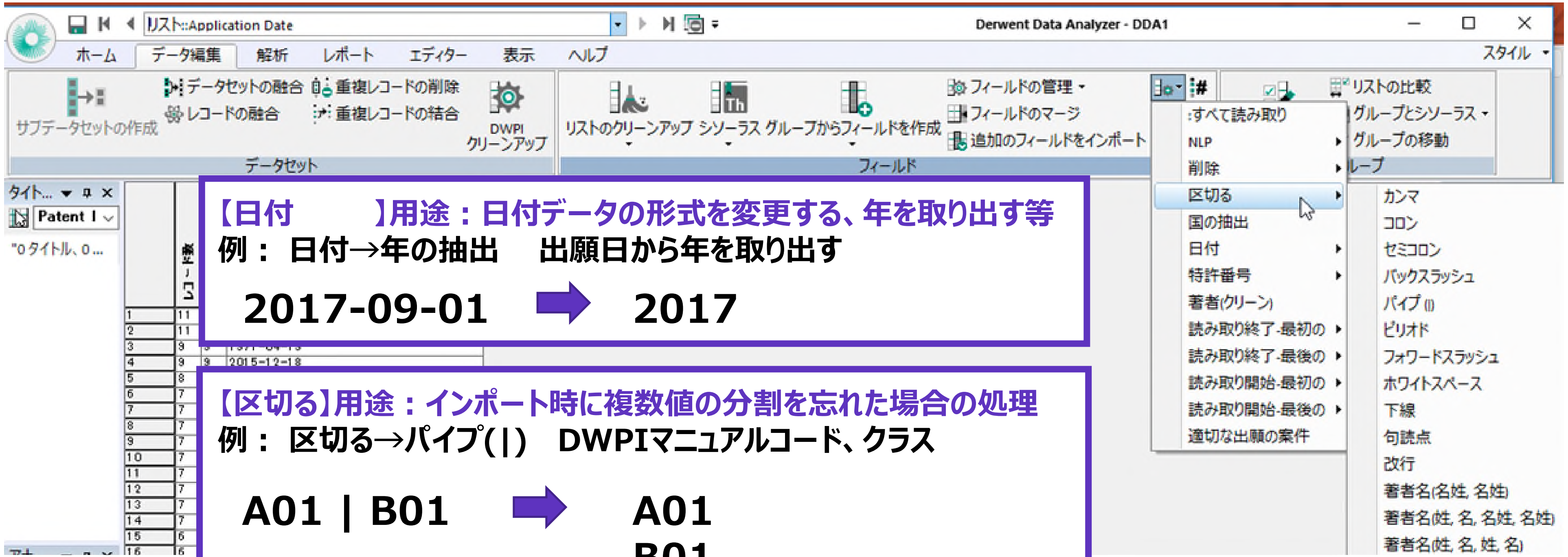
1つのフィールドに複数値が含まれる場合は、**Divide Text** を選択し、**Delimiter**で区切り文字を指定します。
すると、区切られた値が縦に表示されます。
正しく区切られていることを確認してください。

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】



【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

Clarivate™

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

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

【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　列表示

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

2

3 属性を選択して右側に移動させ、OK

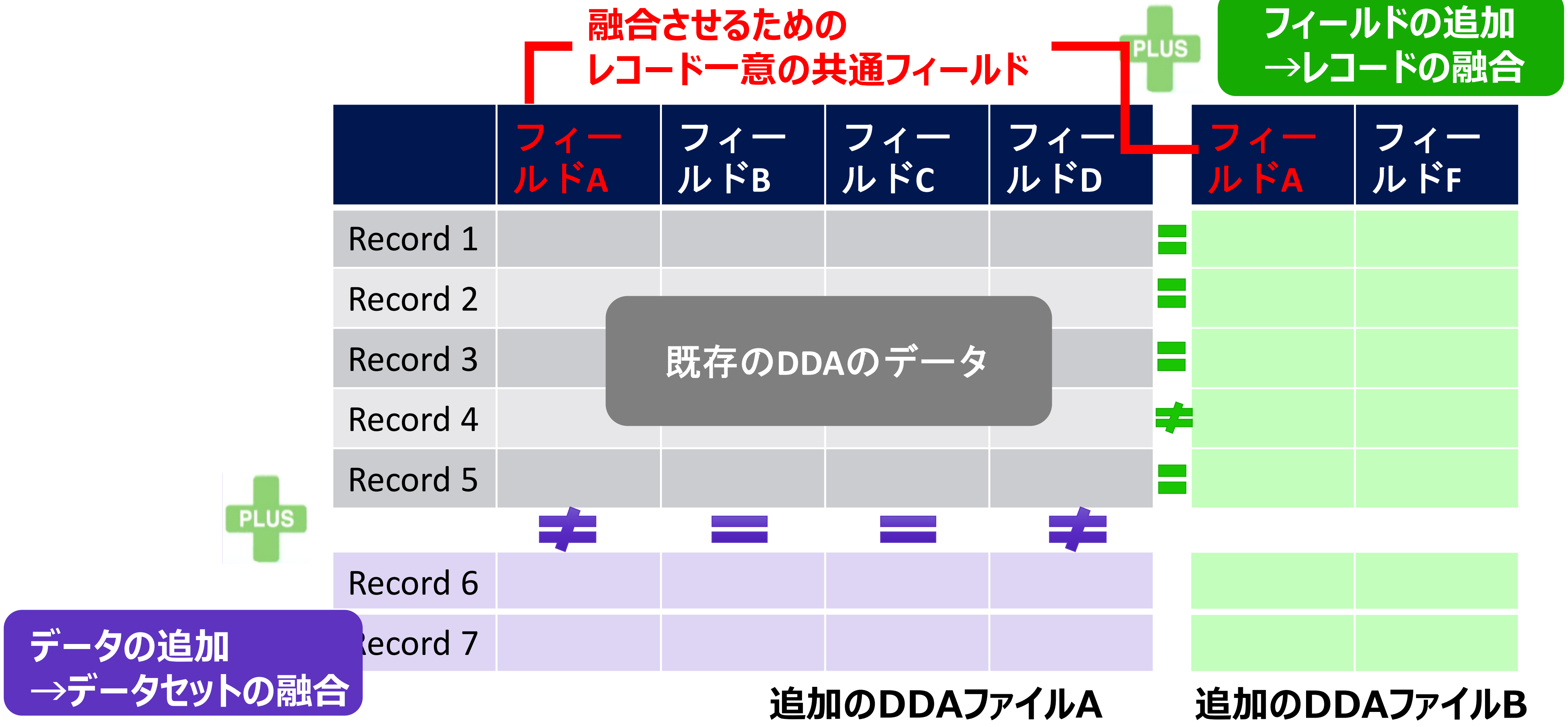
Clarivate™

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。

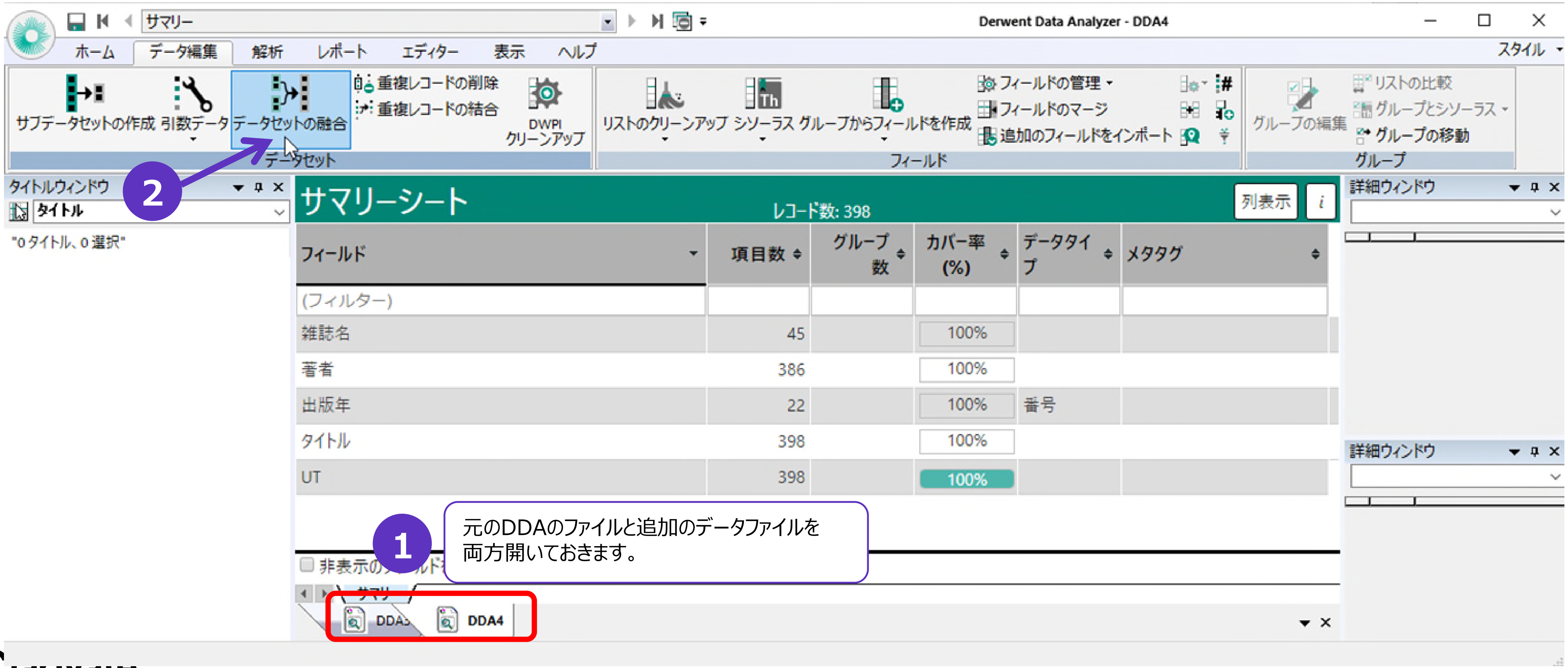


Clarivate™

後からデータを追加する（データセットの融合）

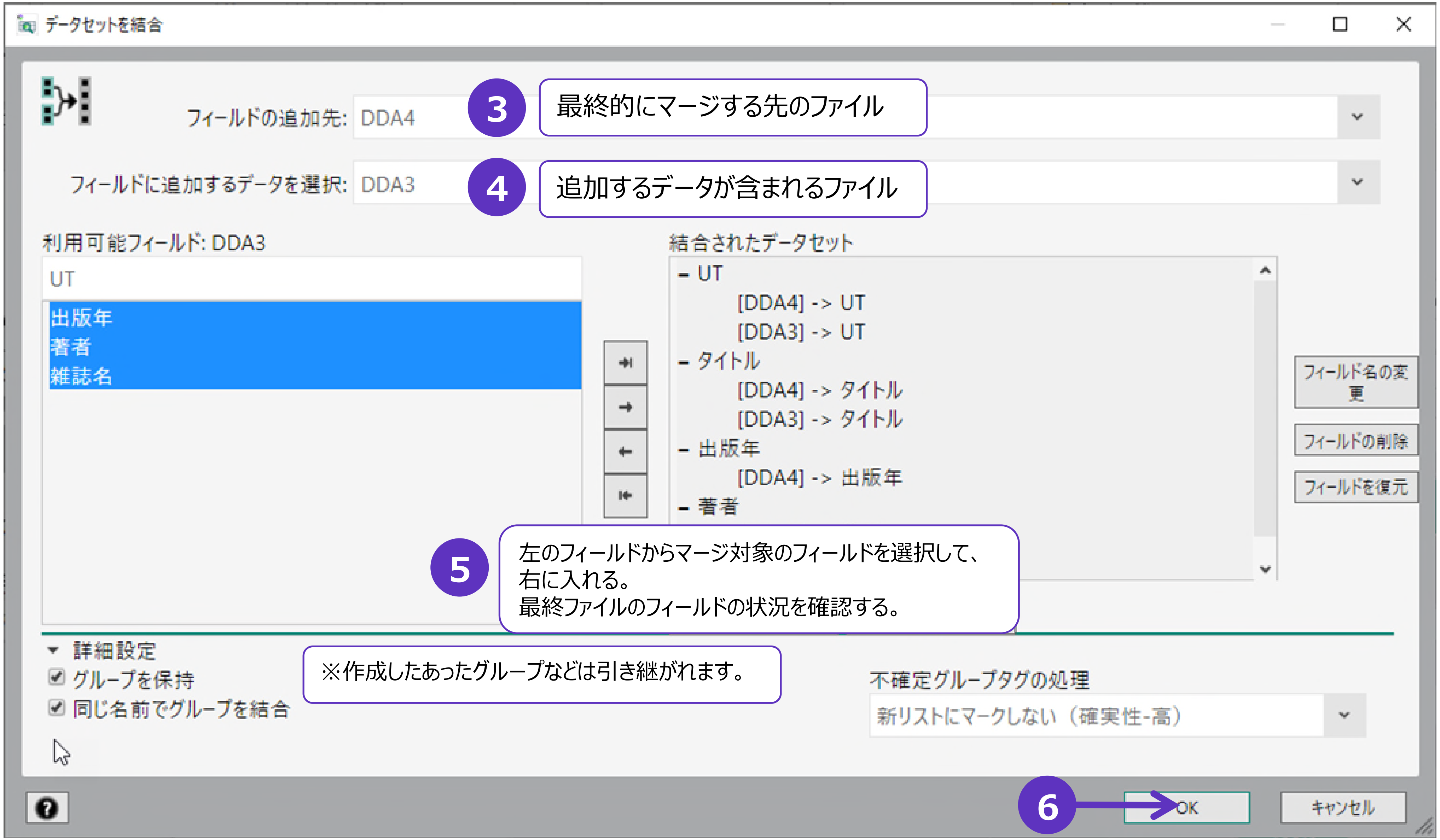
データの追加を行います。同じフィールドを持っている必要はなく、異なるデータソースのデータも取り込めます。
例） Derwent Patent Searchの特許キーワードと文献キーワードを併せて分析する
例） 既存の分析ファイルに新しい年代のデータを追加する
例） 一度にダウンロードできず複数回に分けてダウンロードしたファイルを1つにまとめる

データの追加
→データセットの融合



Clarivate

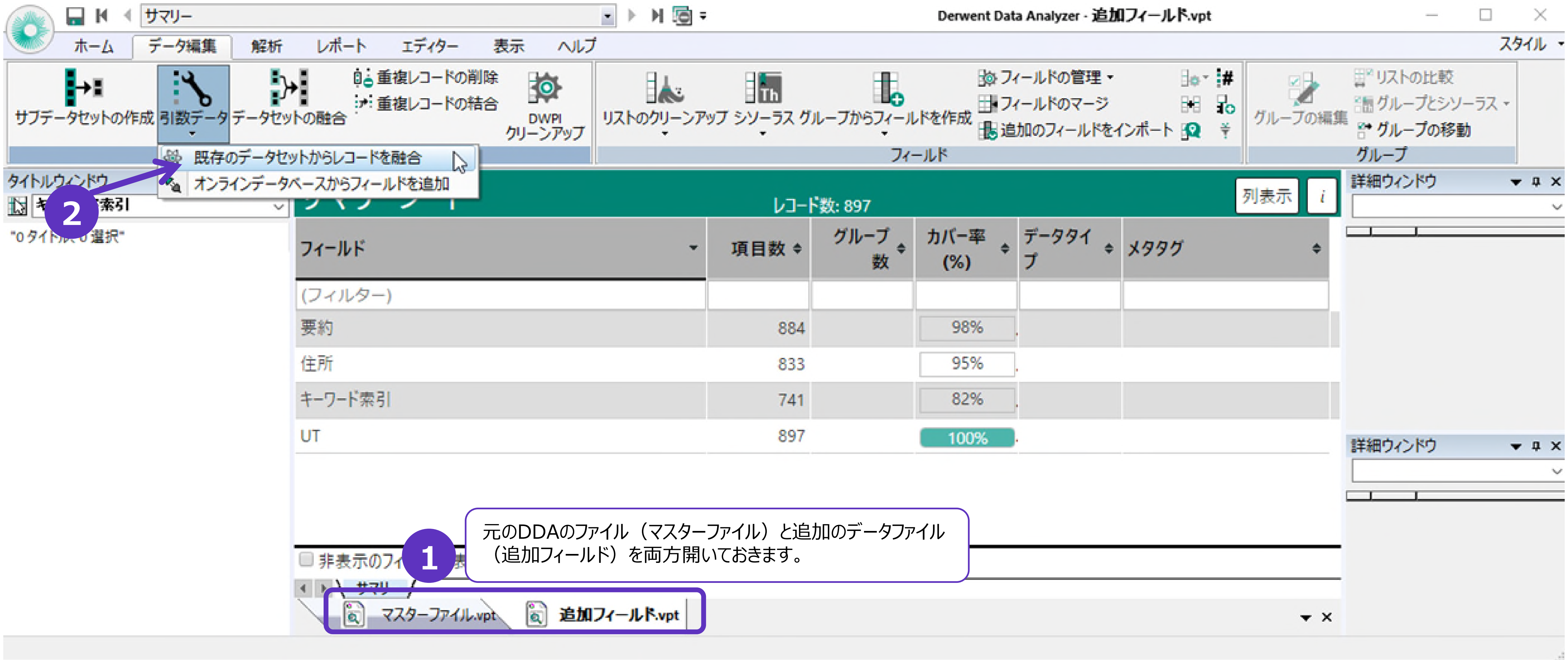
後からデータを追加する（データセットの融合）



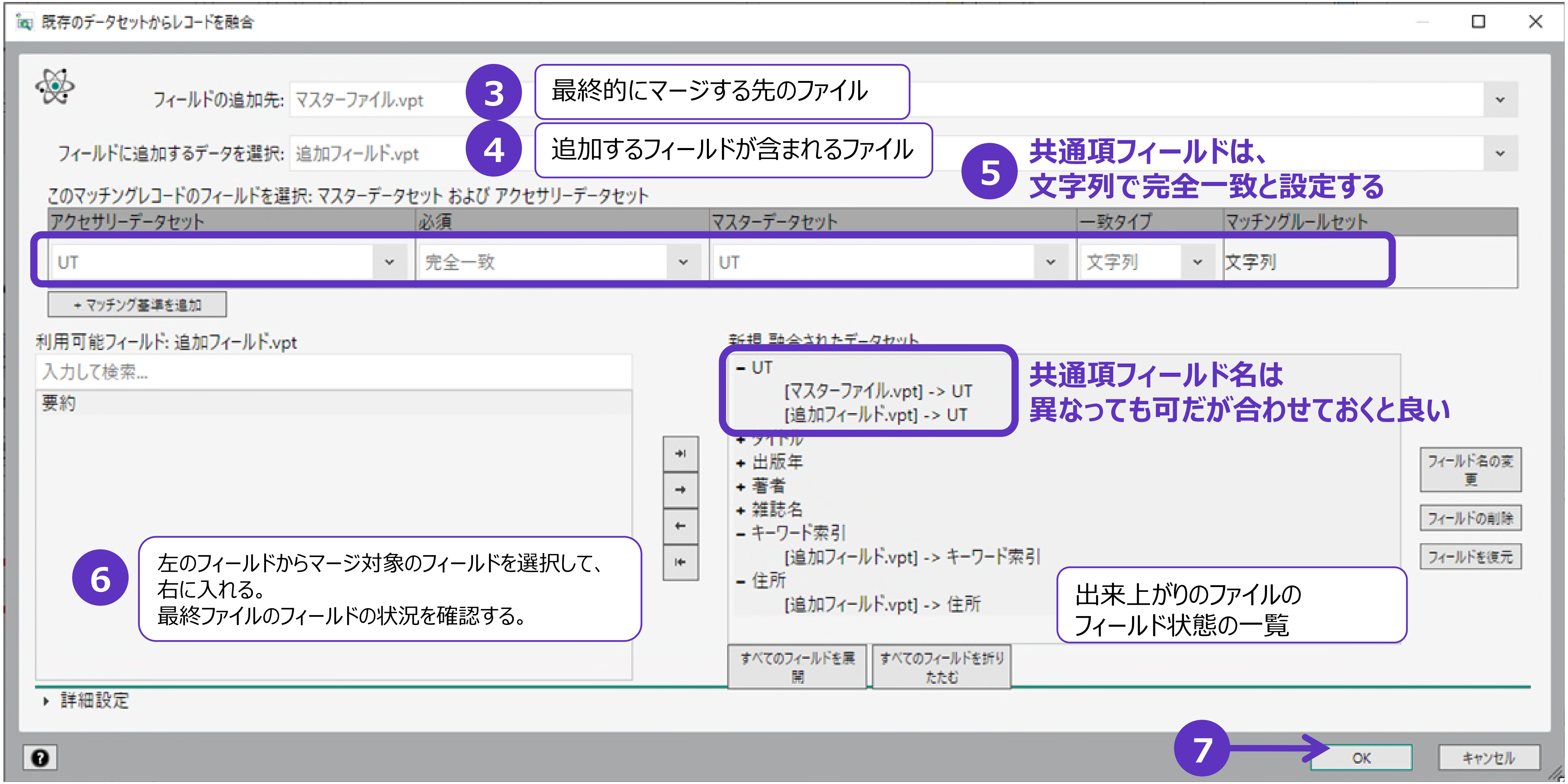
後からフィールドを追加する（レコードの融合）

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 Patent Search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

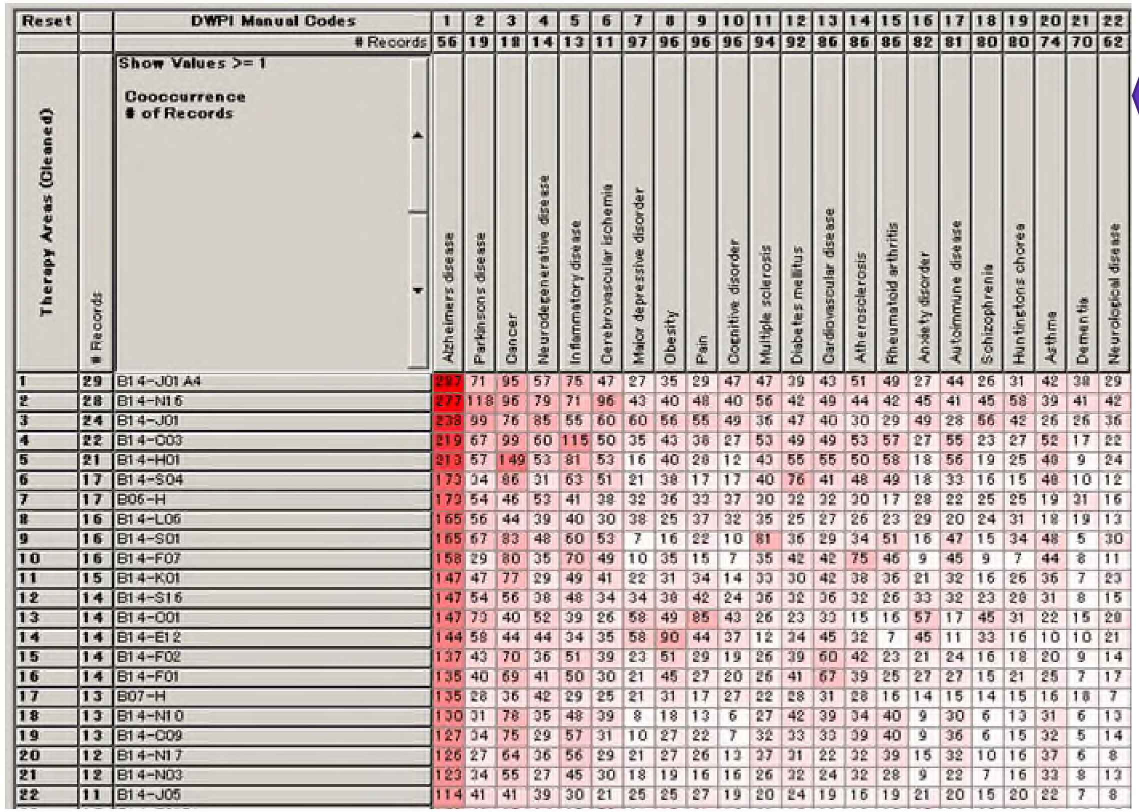
フィールドの追加
→レコードの融合



後からフィールドを追加する（レコードの融合）



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



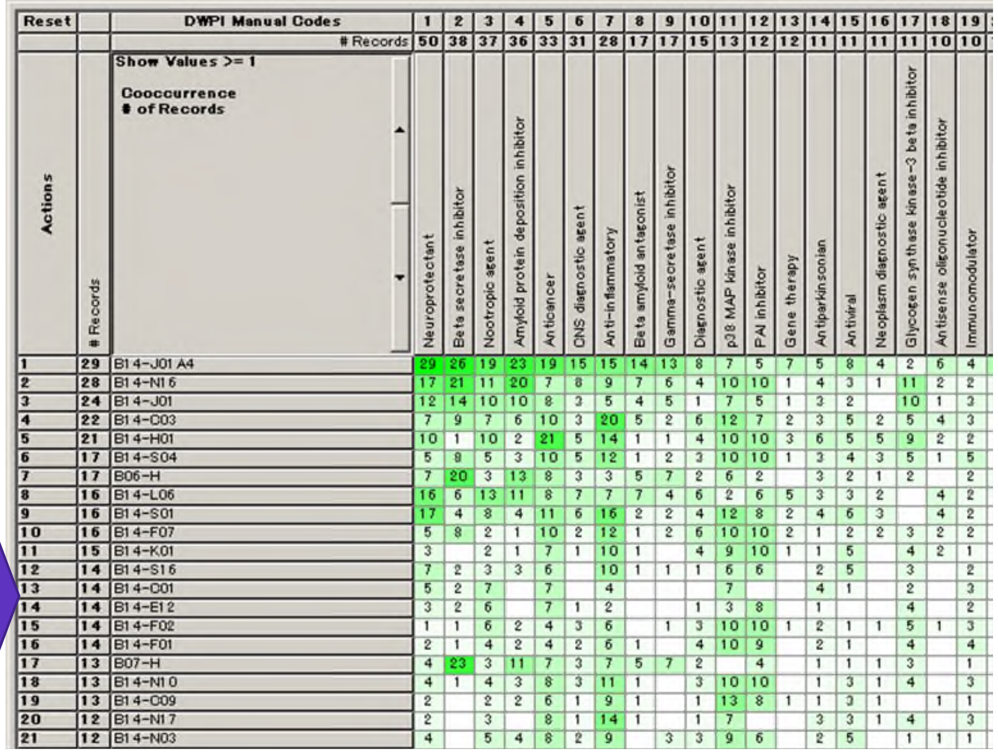
Reset		DWPI Manual Codes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
		# Records	56	19	18	14	13	11	97	96	96	96	94	92	86	86	86	82	81	80	80	74	70	62
Therapy Areas (Cleaned)	# Records	Show Values >= 1 Cooccurrence # of Records	Alzheimers disease	Parkinsons disease	Cancer	Neurodegenerative disease	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	Huntingtons chorea	Asthma	Dementia	Neurological disease
1	29	B14-J01A4	297	71	95	57	75	47	27	35	29	47	47	39	43	51	49	27	44	26	31	42	38	29
2	28	B14-N16	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	172	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	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-N10	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	33	39	40	9	36	6	15	32	5	14
20	12	B14-N17	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 Patent Search=Manual Code x Cortellis = 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

Derwent Patent Search=Manual Code x Cortellis = 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



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
Actions	# Records	Show Values >= 1 Cooccurrence # of Records	Neuroprotectant	Beta secretase inhibitor	Nootropic 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-J01A4	29	26	19	23	19	15	15	14	13	8	7	5	7	5	8	4	2	6	4
2	28	B14-N16	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	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	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	3
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-N10	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	1
20	12	B14-N17	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	1

Clarivate™

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。

Clarivate™

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、機械学習をどこ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

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。

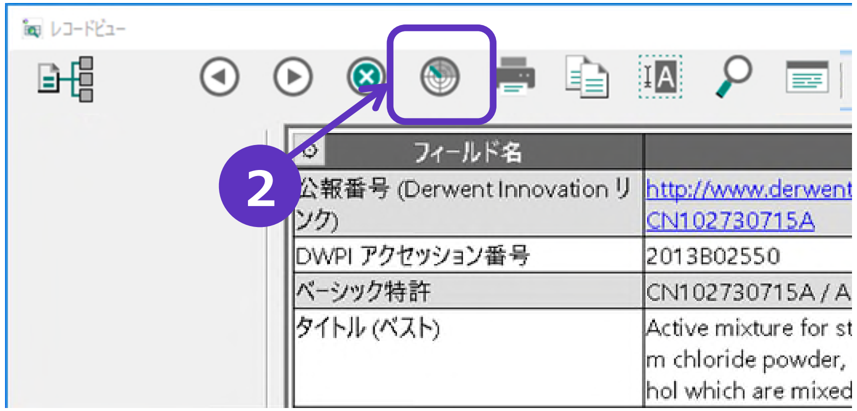
フィールドで、右クリックして、**「メタタグを設定」**をクリックすると、設定できます。

Clarivate™

DDA集合内に存在するレコードの内容に類似する他のレコードを探す

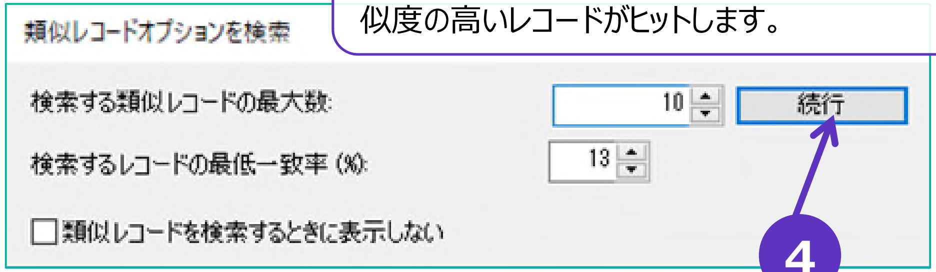
ターゲットが1レコードの場合

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

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。

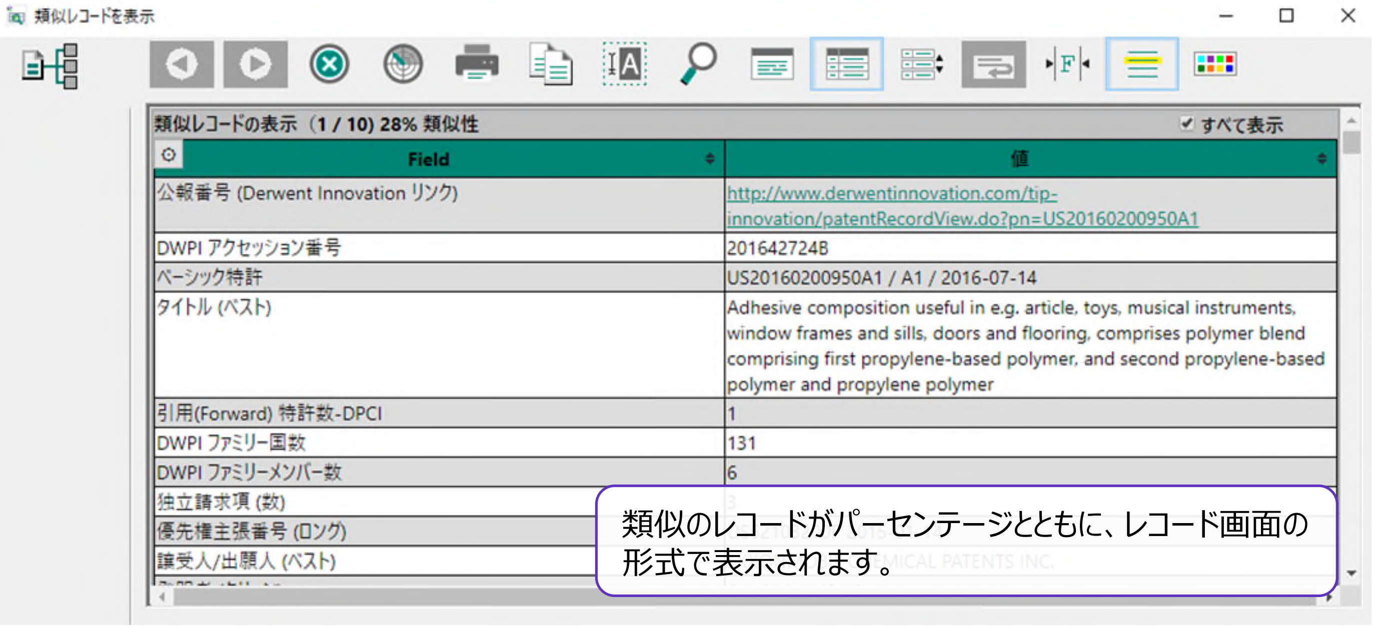


4

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。

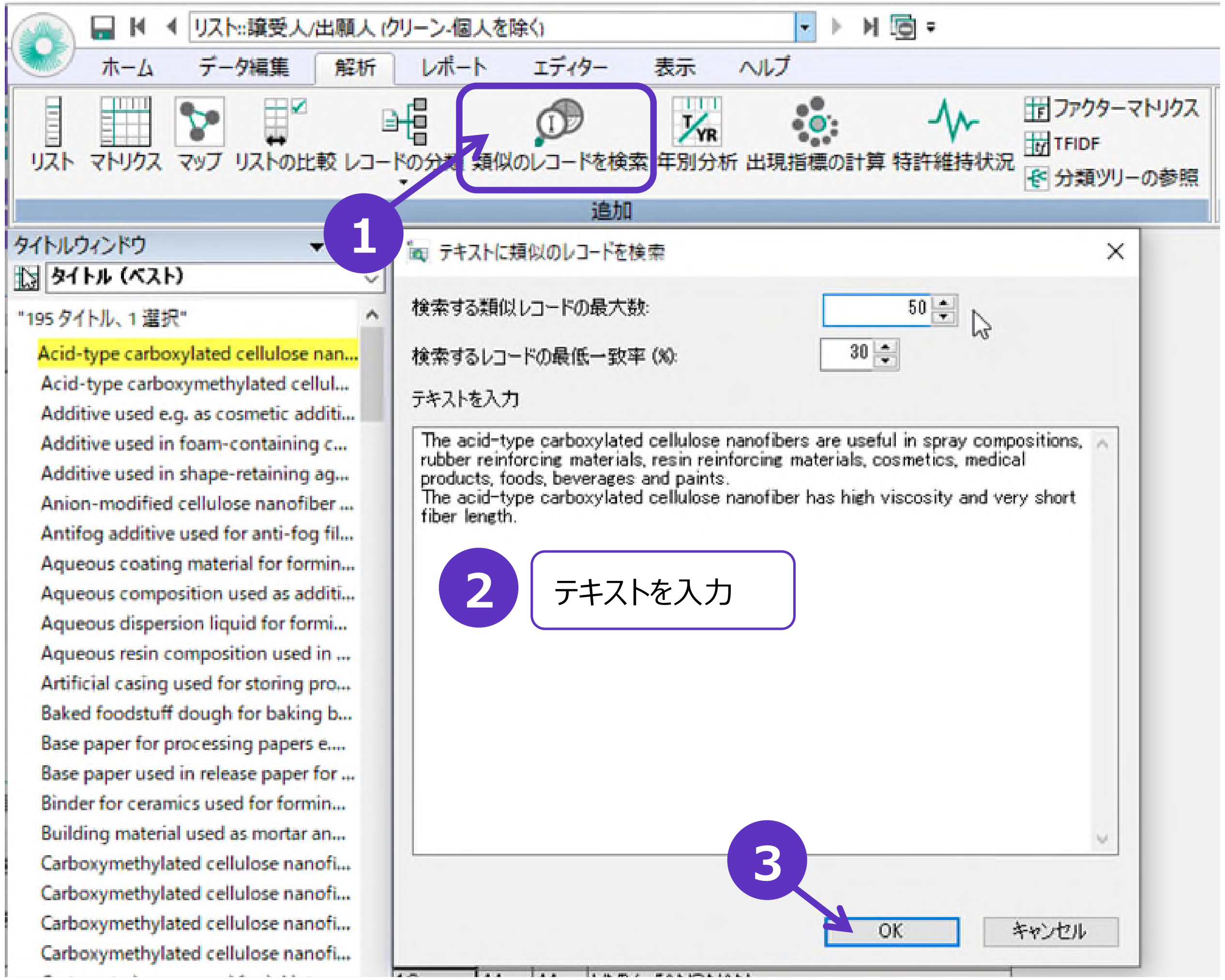
2 右クリックして「類似レコードを検索」をクリック。

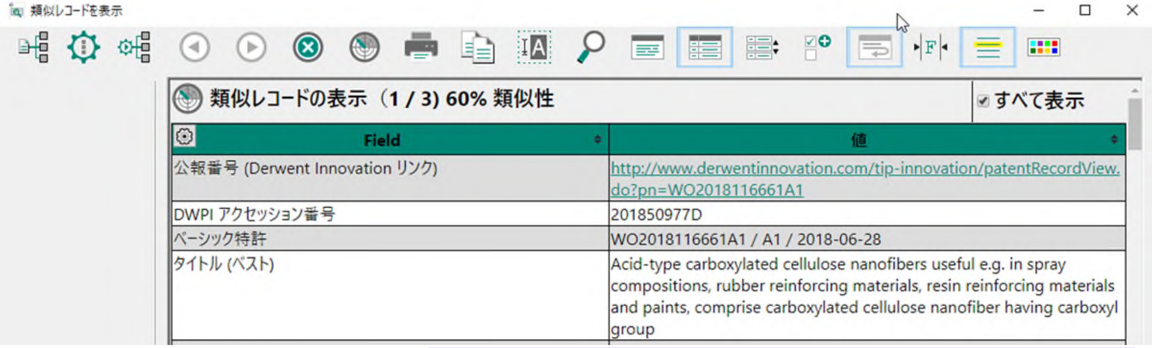


類似のレコードがパーセンテージとともに、レコード画面の形式で表示されます。

Clarivate™

DDA集合内に存在しない内容（任意のテキスト）に類似するレコードを探す





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。

Clarivate™

機械学習機能を利用して自動分類を行う

Clarivate™

Derwent Data Analyzer 自動分類の仕組み

学習

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

付与

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

カテゴリ A
抄録に cookie, cake

カテゴリ A
抄録に cookie, desert

カテゴリ B
抄録に drink, beverage

カテゴリ B
抄録に drink, alcohole

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）

分類A
- 特許1: a c d / e h i / m o q
- 特許2: a b e / f g j / k x w

分類B
- 特許3: a b c / d f g / k n i
- 特許4: b e g / h m j / o s t

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

分類A
- 特許1: a c d / e h i / m o q
- 特許2: a b e / f g j / k x w

分類B
- 特許3: a b c / d f g / k n i
- 特許4: b e g / h m j / o s t

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
- 特許5: b a c / e i j / x p q
- 特許6: a b f / g k m / s x w
- 特許7: d a b / f e n / y x z
- 特許8: c b e / f g h / i s t

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

分類A
- 特許5: b a c / e i j / x p q
- 特許7: d a b / f e n / y x z

分類B
- 特許6: a b f / g k m / s x w
- 特許8: c b e / f g h / i s t

Clarivate™

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…)

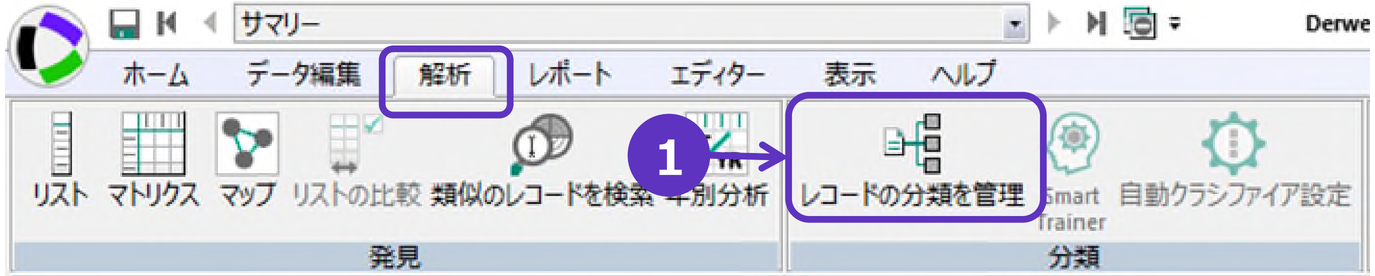
↓

４．残りのレコードに対して自動で分類を付与する

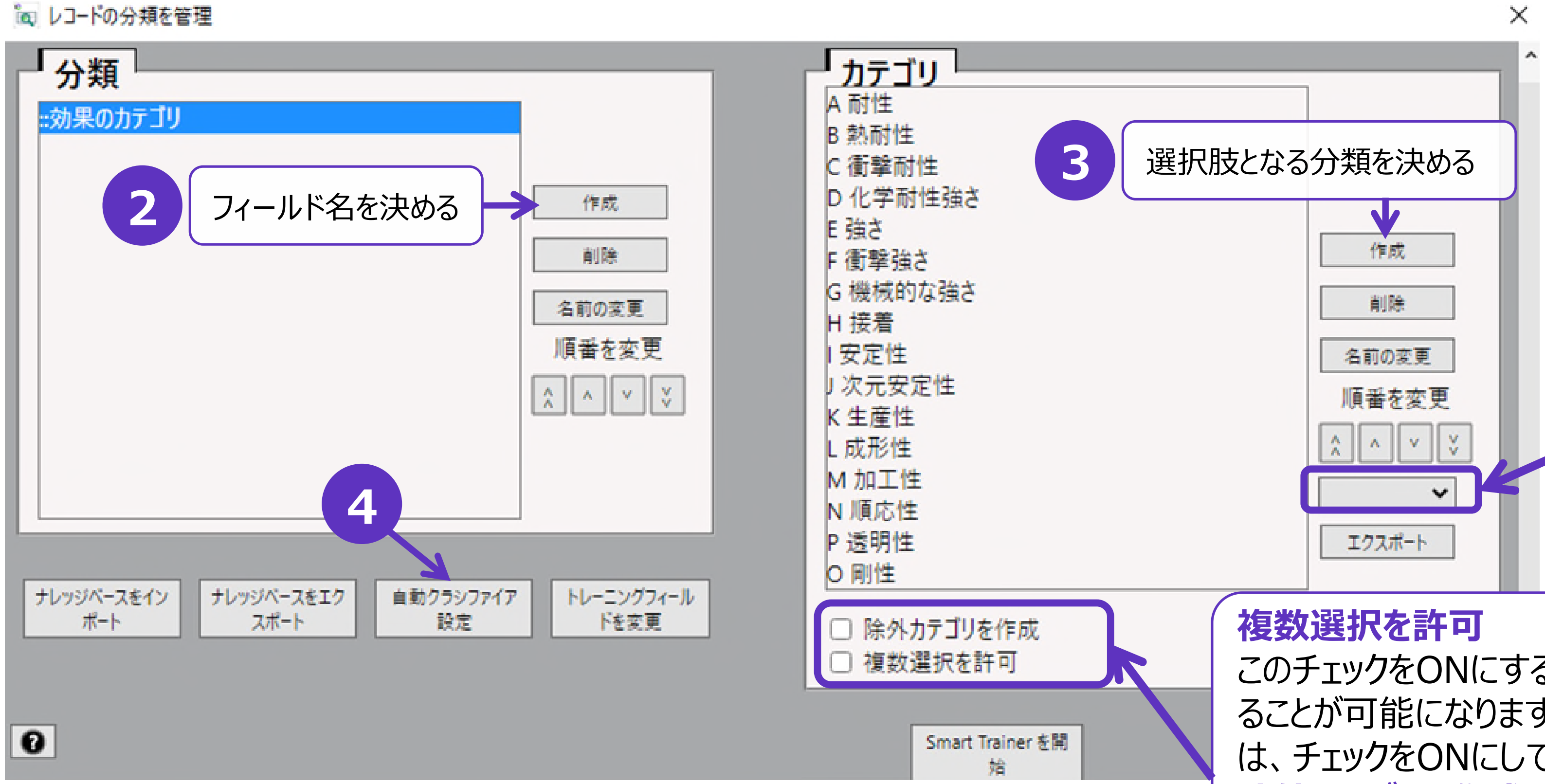
Clarivate™

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



1



2 フィールド名を決める

3 選択肢となる分類を決める

4

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）

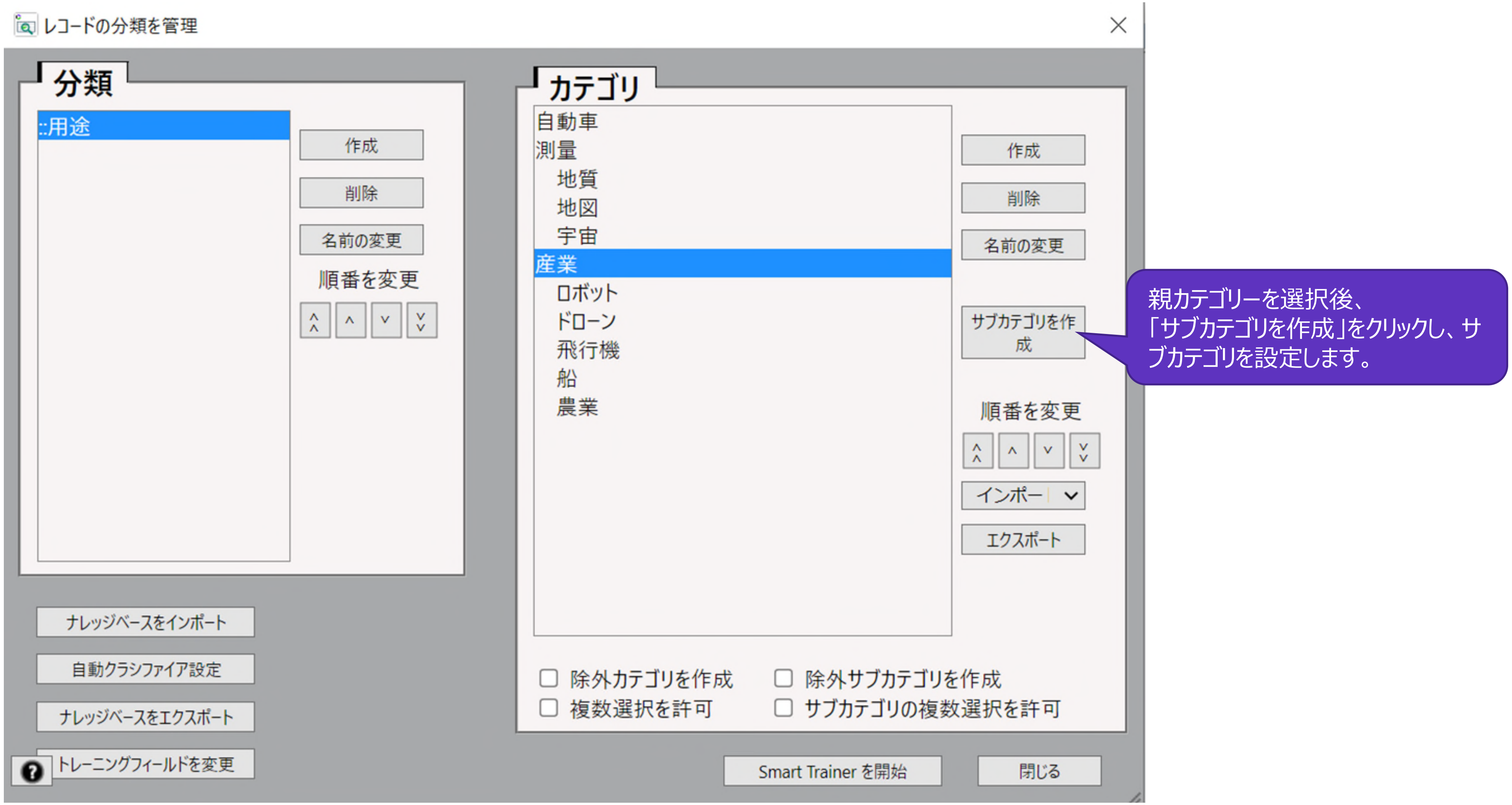
複数選択を許可
このチェックをONにすると、1つのレコードに対し、複数の分類を付与することが可能になります。複数の分類に当てはまる可能性がある場合は、チェックをONにしておきます。

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™

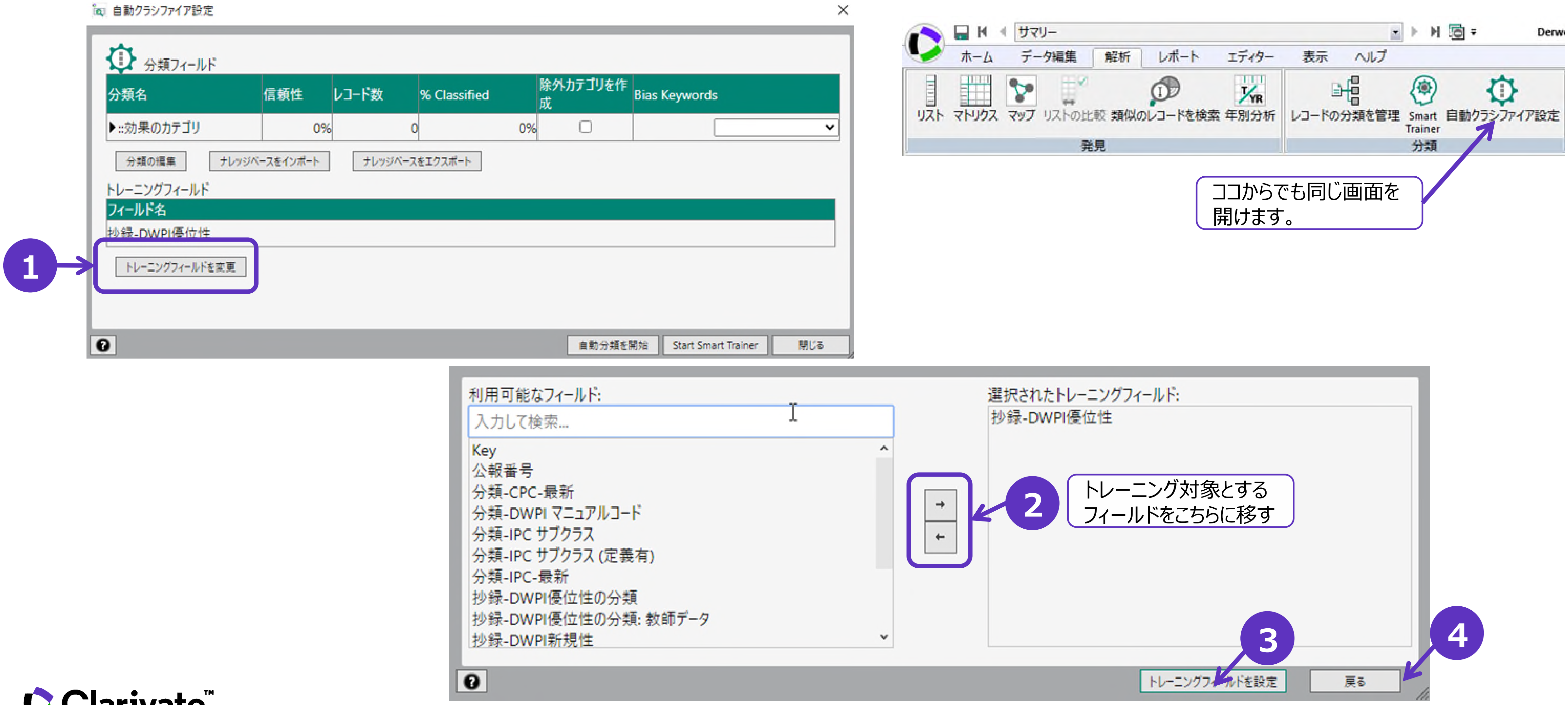
参考：分類のサブカテゴリ

◆ レコード分類の各カテゴリーにサブカテゴリー（下の階層のカテゴリー）を作成できます。サブカテゴリーに分類されたレコードは親のカテゴリー（上の階層のカテゴリー）にも分類されます。



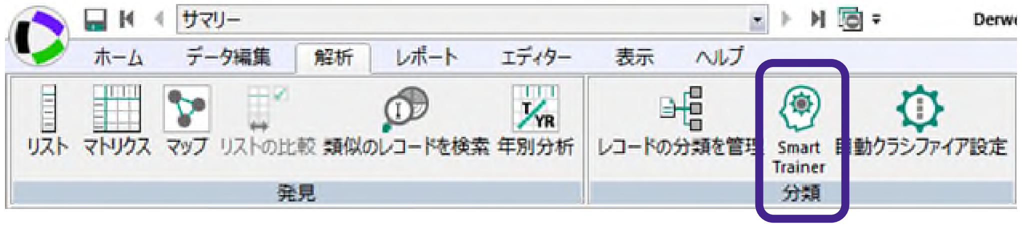
DDAで教師データを作成して、自動分類を行う

2．トレーニングフィールドの設定

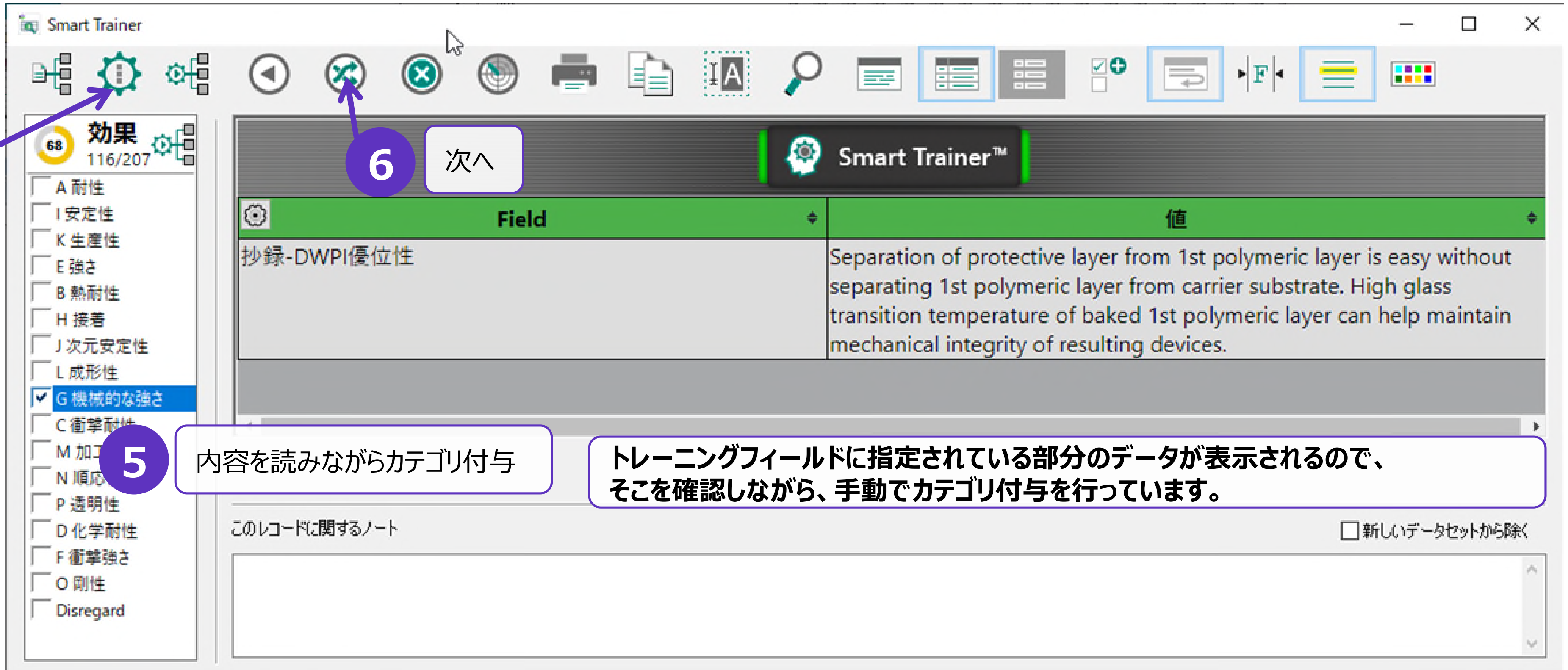


DDAで教師データを作成して、自動分類を行う
3．レコード画面で手動でカテゴリを付与（教師データの作成）

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

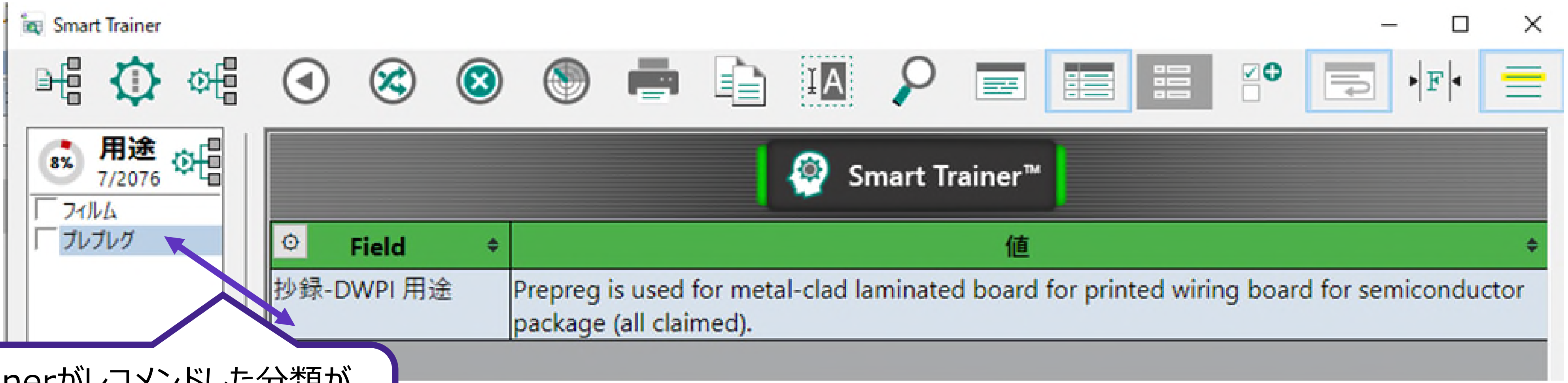
5 内容を読みながらカテゴリ付与

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。

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

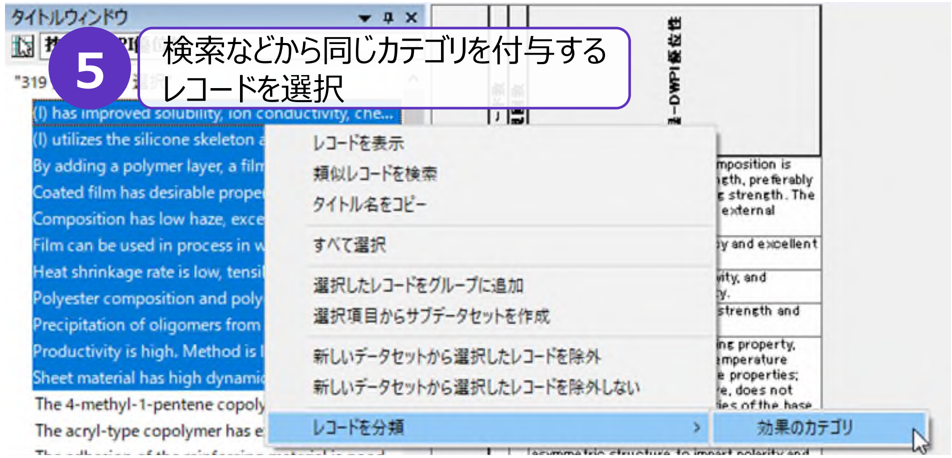
3．レコード画面で手動でカテゴリを付与（教師データの作成）

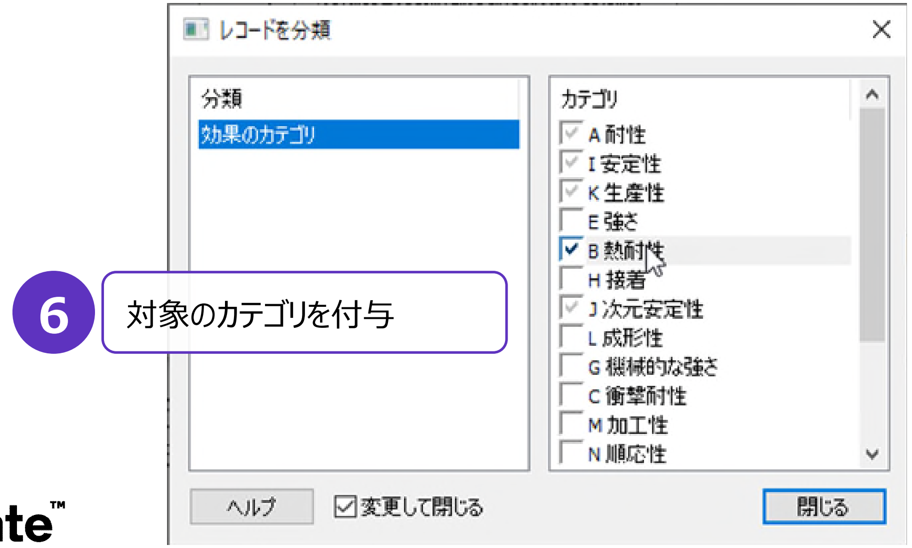
B. レコード画面を使って行う

ユーザーの手作業をDDAが機械学習していきます・・・・

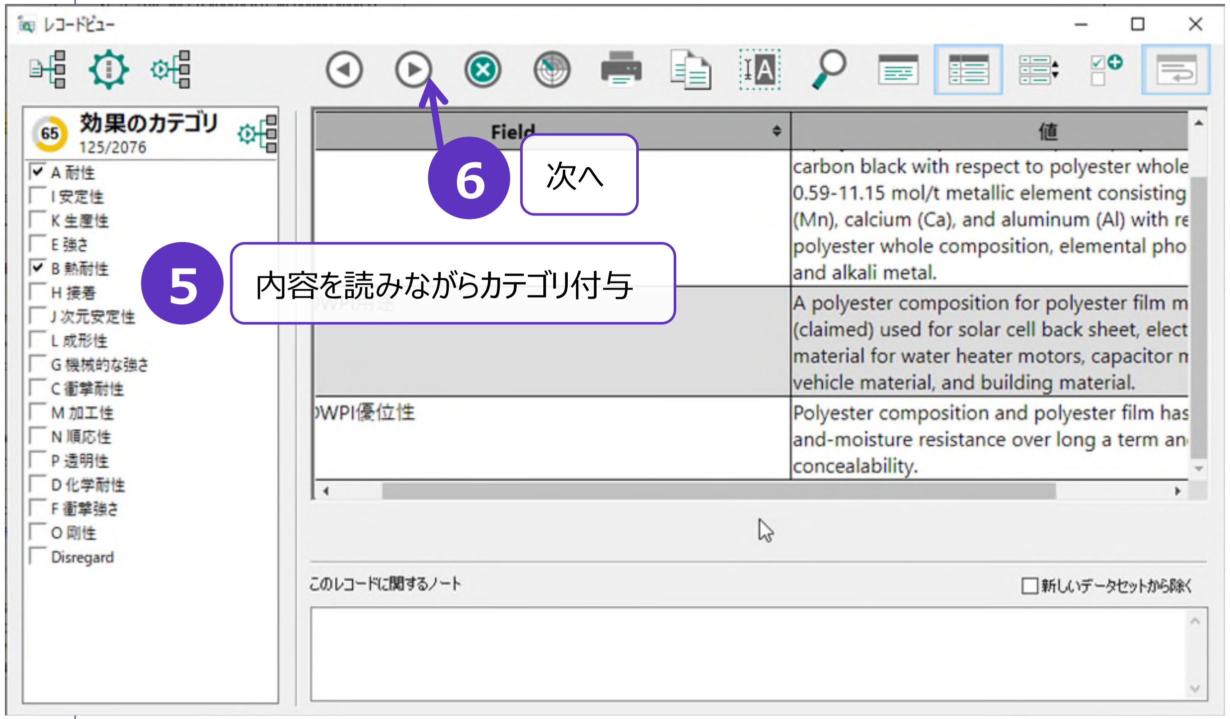
ユーザーによる分類の付与が始まると、その内容にしたがい、自動的に機械学習が始まります。付与を行えば行うほど、信頼度が向上します。

【複数に一気に付与】





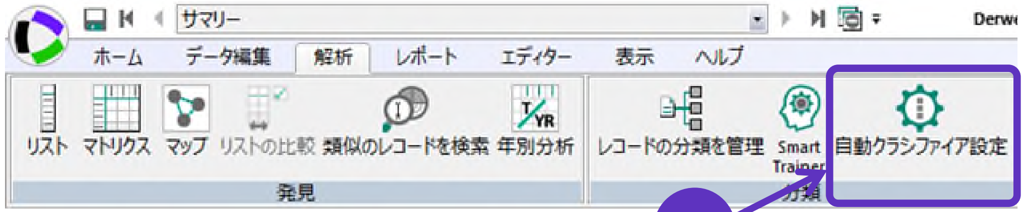
【レコード画面で1レコードずつ付与】



Clarivate™

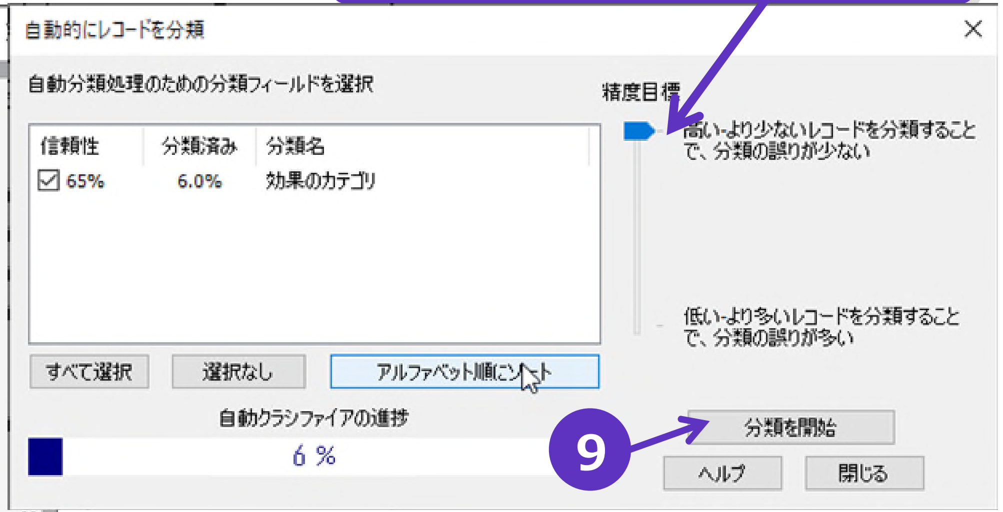
DDAで教師データを作成して、自動分類を行う

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。
精度を高めにすることをお勧めします。



参考：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」ー「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。

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

↓

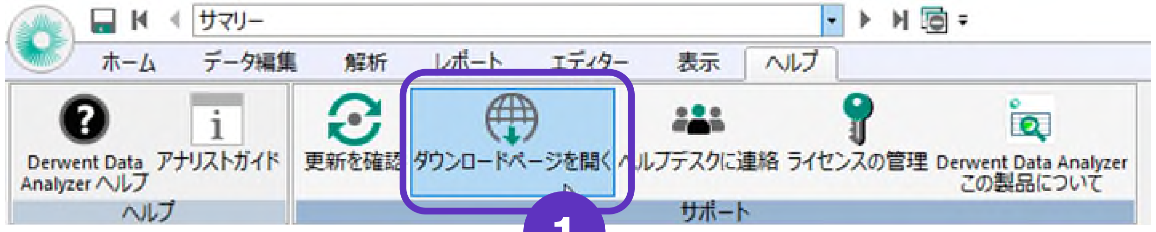
6．未分類データへ自動で分類を付与する

↓

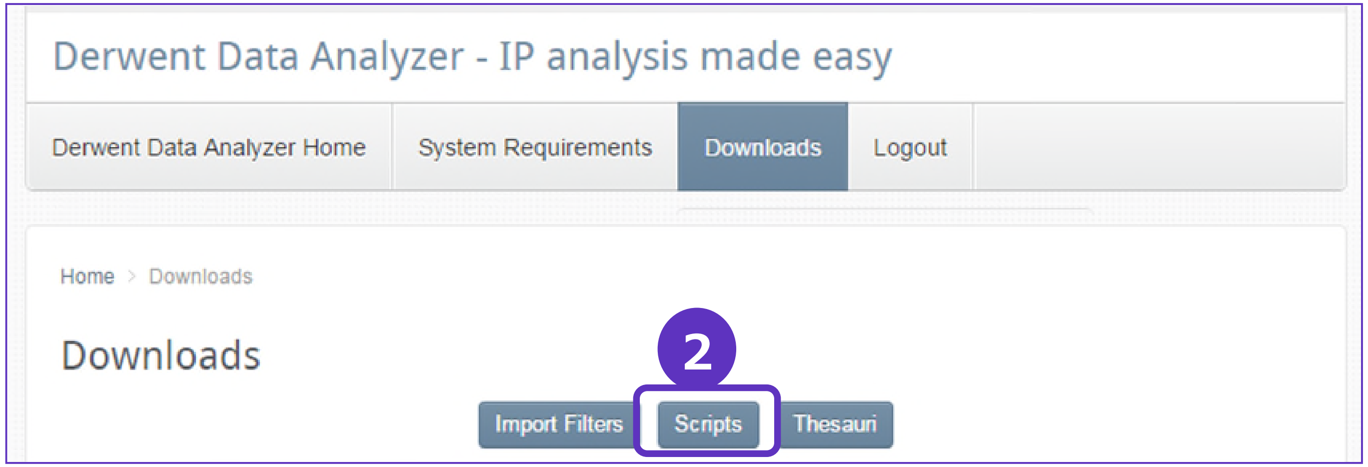
7．必要に応じて自動付与データをダウンロードまたは再分類

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。

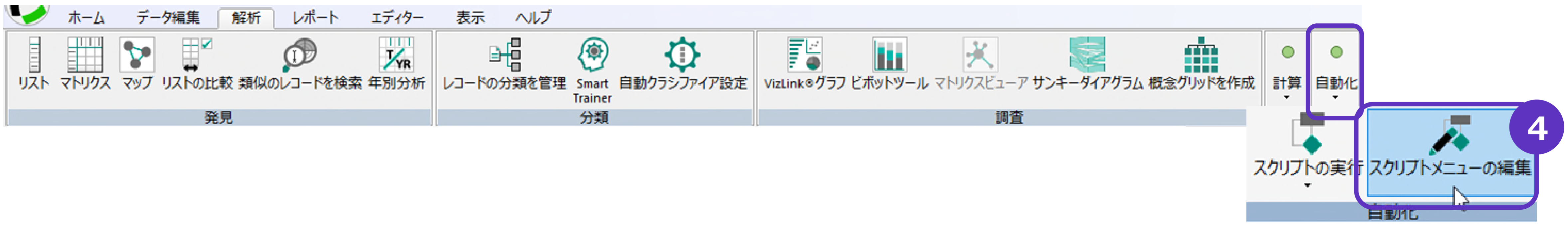


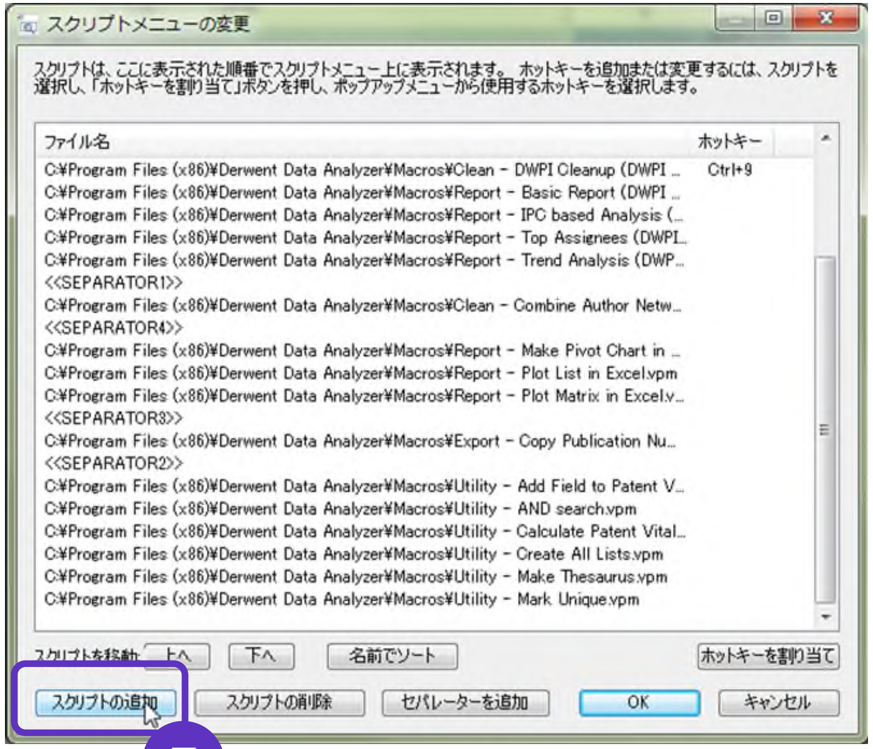
ダウンロードページが表示されます。

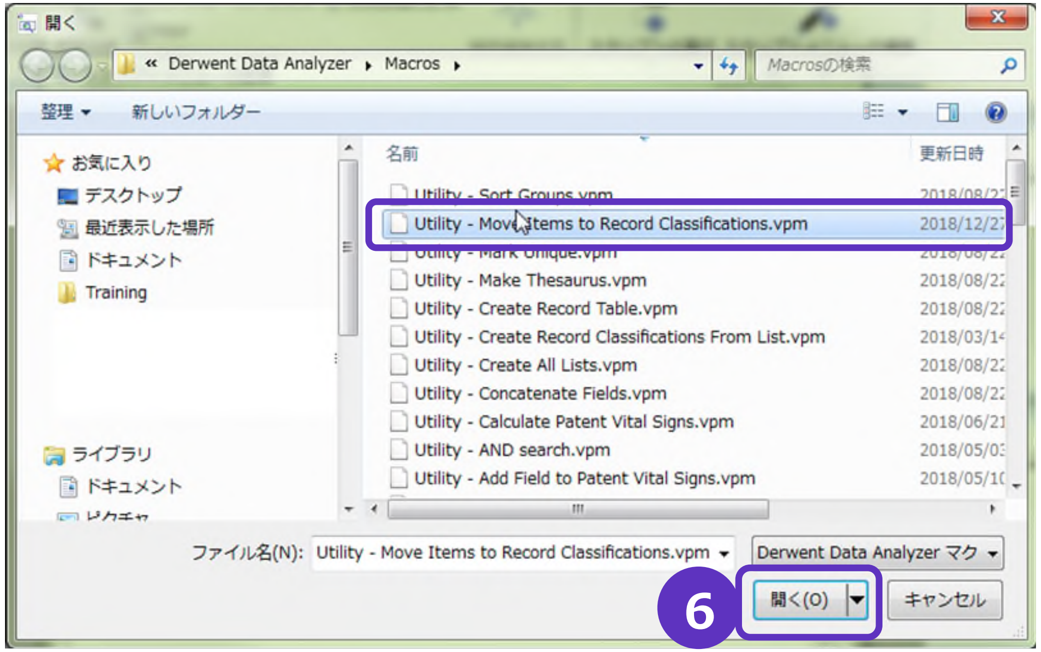
3

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

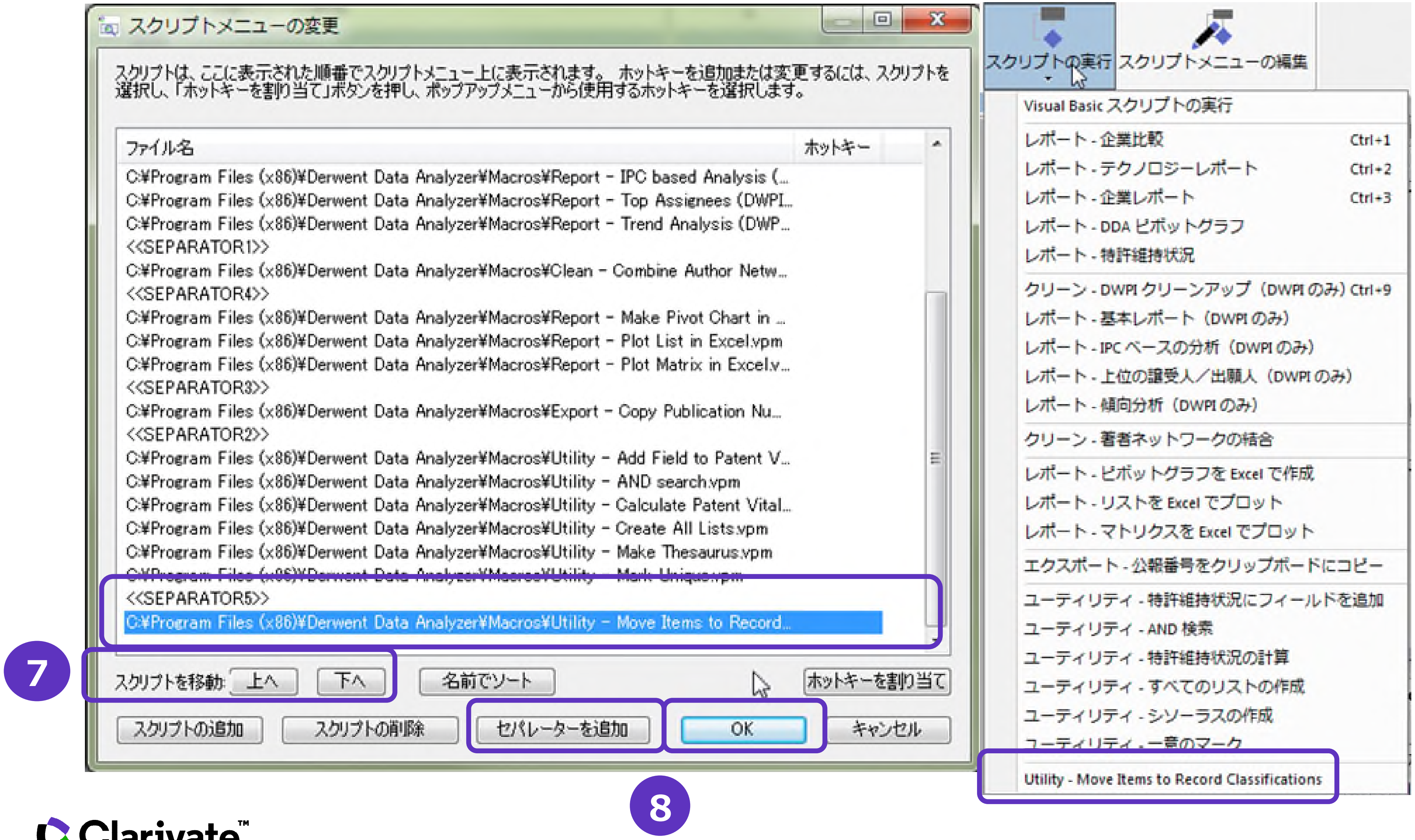






先ほどダウンロードした
Utility – MoveItems to Record Classifications.vpm
を指定して「開く」

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。



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。

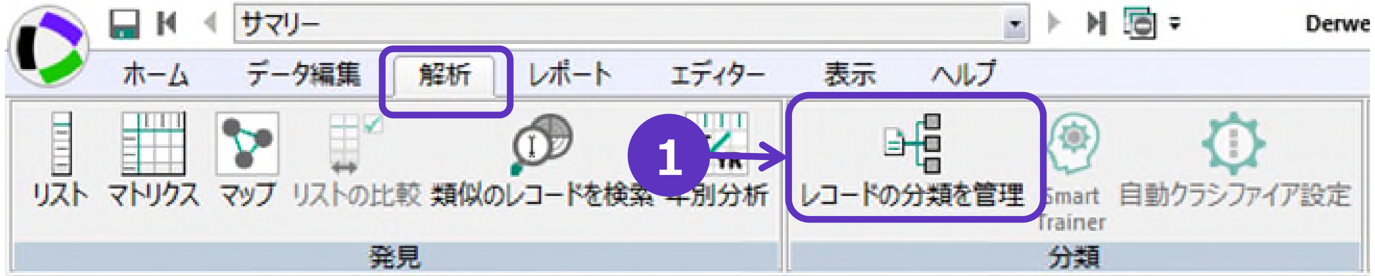
教師データとなる情報と未分類のデータを１つのファイルにしておくとインポートが一回で済みます。
（※教師データと未分類のデータでファイルを分けてある場合、別々に取り込んだ後、マージする方法でも行うことができます (データセットの融合)。）

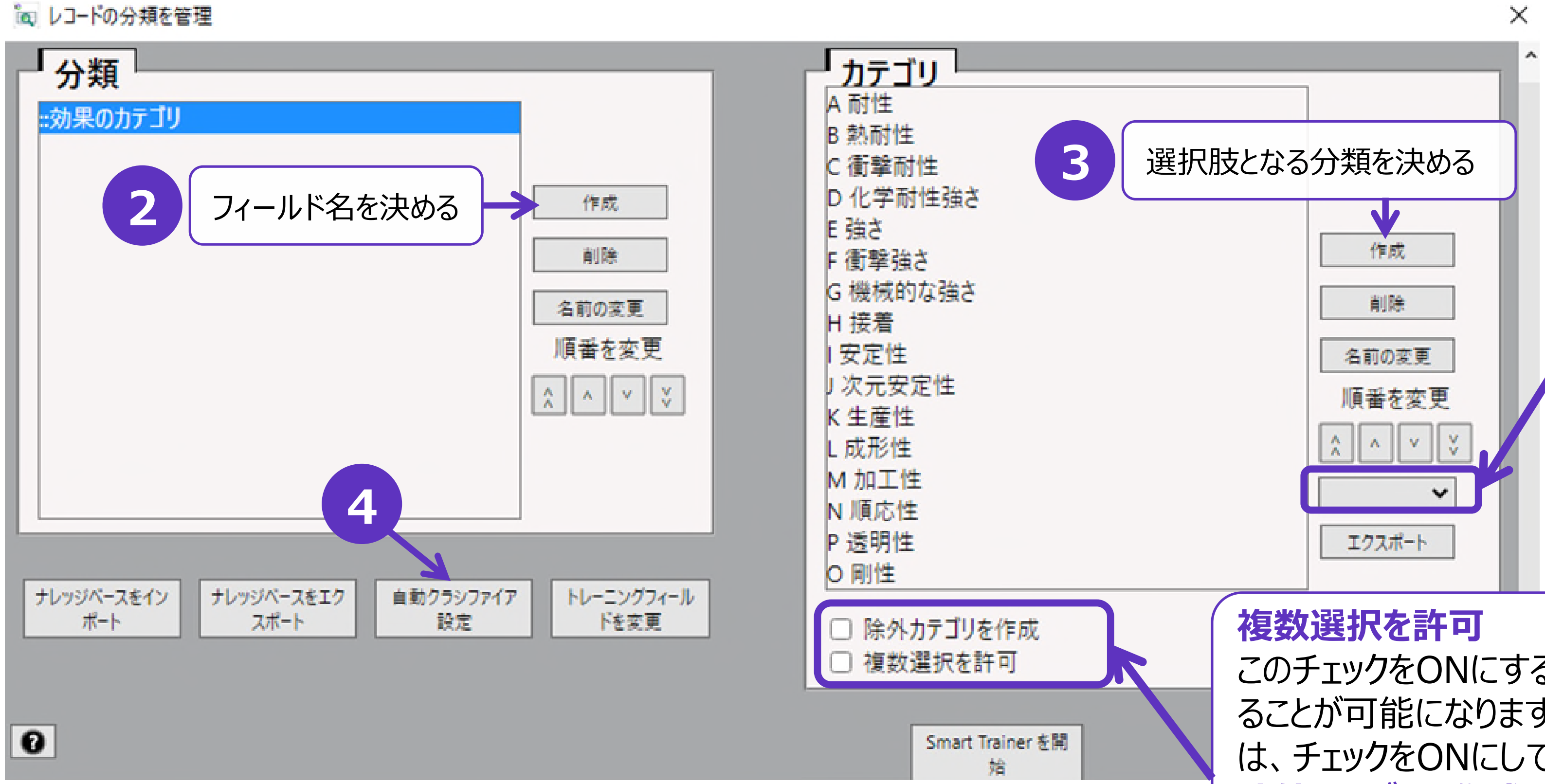
分類済みの教師データについては、
既存の分類を入力した状態にします。

	A	B	C	D	E	F	G
1	公報番号	抄録-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 terep	A フィルム; H 樹脂	Polybutylene terephthala	I 安定性; K 生産性; J 次元安定性	A polybutylene terephtha	D フィルム; F 建材; K 包装
3	WO2013105340A1	A carbon fiber reinfor	D カーボン; G 繊維; H 樹脂	The carbon fiber has large	A 耐性; E 強さ; C 衝撃耐性	Carbon fiber reinforced p	A 自動車; E モバイル; H 飛行機
4	WO2013105300A1	Aromatic polyamide p	A フィルム; J ポリアミド	The short circuit of the pc	A 耐性; B 熱耐性	Aromatic-polyamide porc	A 自動車
5	KR2013081453A	A white opaque polye	A フィルム; B ポリエステル; F 組成物; H 樹脂	The white opaque polyes	E 強さ	White opaque polyester f	D フィルム
6	KR2013081124A	A heat-shrinkable pol	A フィルム; B ポリエステル; H 樹脂	The heat-shrinkable poly	A 耐性	Heat-shrinkable polyeste	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 u	I 安定性; K 生産性; J 次元安定性	The polyimide film is use	D フィルム
8	KR2013076391A	Manufacture of meta-	A フィルム	The meta-aramid-based f	E 強さ	Manufacture of meta-ara	C 製造; D フィルム
9	KR2013076390A	Meta-aramid film con	A フィルム	The film exhibits exceller	E 強さ; G 機械的な強さ	The meta-aramid film is u	D フィルム
10	KR2013076044A	An adhesive protectic	A フィルム; F 組成物; H 樹脂	The adhesive protection f	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 lam	D フィルム
12	FR2985263A1	Recycling a multilayer	A フィルム	The process: produces re	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 comprise	A フィルム; B ポリエステル	The process enables effic	A 耐性; I 安定性; K 生産性; J 次元	The process is useful for f	D フィルム
14	WO2013099741A1	A carbon fiber base m	D カーボン; G 繊維	The carbon fiber base ma	E 強さ	Carbon fiber base materia	A 自動車; B プレプレグ; H 飛行
15	WO2013099608A1	A biaxially-oriented p	A フィルム; B ポリエステル	The biaxially-oriented po	A 耐性; I 安定性; K 生産性; L 成形	Biaxially-oriented polyes	A 自動車; D フィルム; E モバイ
16	WO2013099234A1	A polyphenylene sulf	F 組成物; H 樹脂	The polyphenylene sulfic	A 耐性; I 安定性; K 生産性; L 成形	Polyphenylene sulfide re	A 自動車
17	WO2013099707A1	A sizing agent coated	D カーボン; G 繊維	The sizing agent coated ca	A 耐性; H 接着; I 安定性; K 生産	Sizing agent coated carbo	A 自動車; B プレプレグ; G 船舶
18	WO2013099862A1	An epoxy resin compo	C エポキシ樹脂; F 組成物; H 樹脂	The epoxy resin composit	A 耐性; E 強さ; B 熱耐性	Epoxy resin composition f	A 自動車; B プレプレグ; F 建材
19	WO2013099111A1	A thermoplastic resin	B ポリエステル; F 組成物; H 樹脂	The molded article obtair	A 耐性; I 安定性; B 熱耐性	Thermoplastic resin composition is used for laser-marking	
20	WO2013099607A1	A polyolefin micropor	A フィルム; H 樹脂	The polyolefin microporo	A 耐性; E 強さ	Polyolefin microporous fi	D フィルム
21	WO2013099539A1	A polyolefin micropor	A フィルム	The polyolefin microporo	K 生産性	Polyolefin microporous fi	D フィルム; E モバイル; J コン

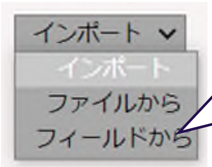
既存の分類データを利用 (教師データの作成はDDA外で行う)して、自動分類を行う

3．分類フィールドの作成





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）



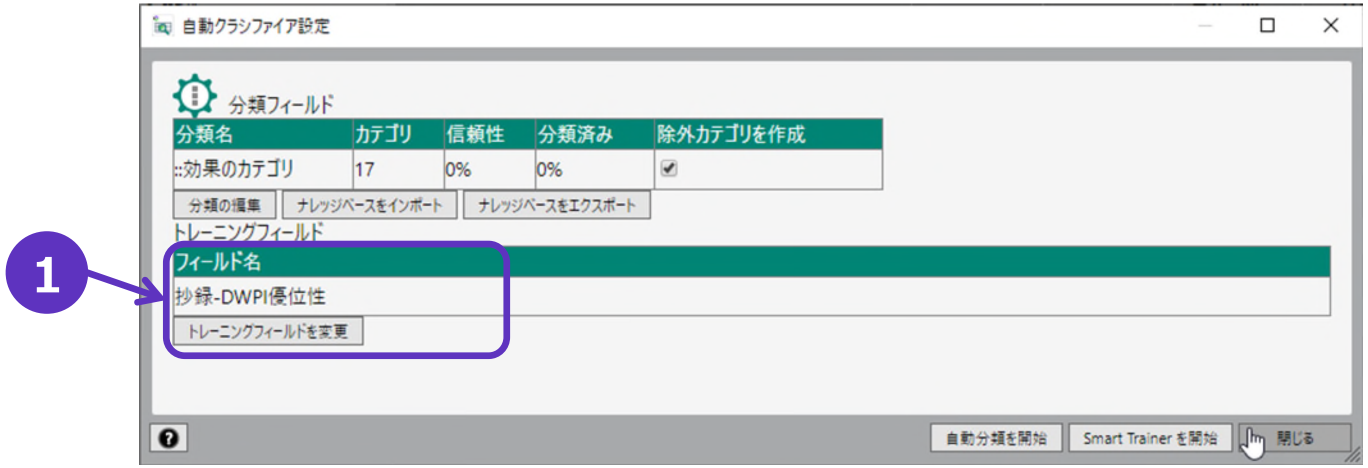
教師データの分類フィールドの項目を活用する場合は、「フィールドから」を選択し該当のフィールドをインポート

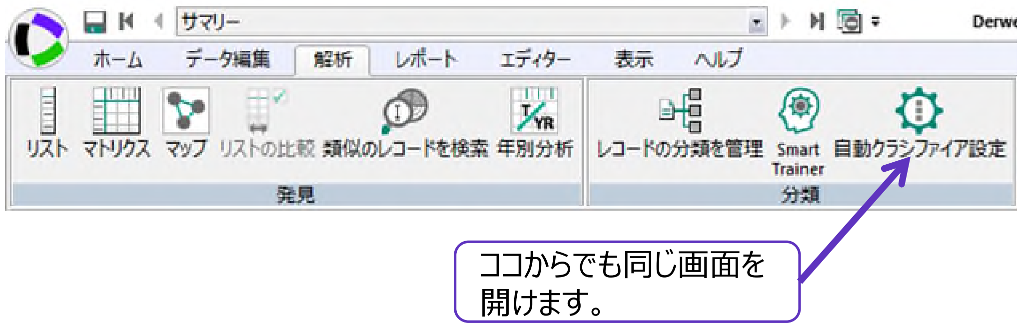
複数選択を許可
このチェックをONにすると、1つのレコードに対し、複数の分類を付与することが可能になります。複数の分類に当てはまる可能性がある場合は、チェックをONにしておきます。

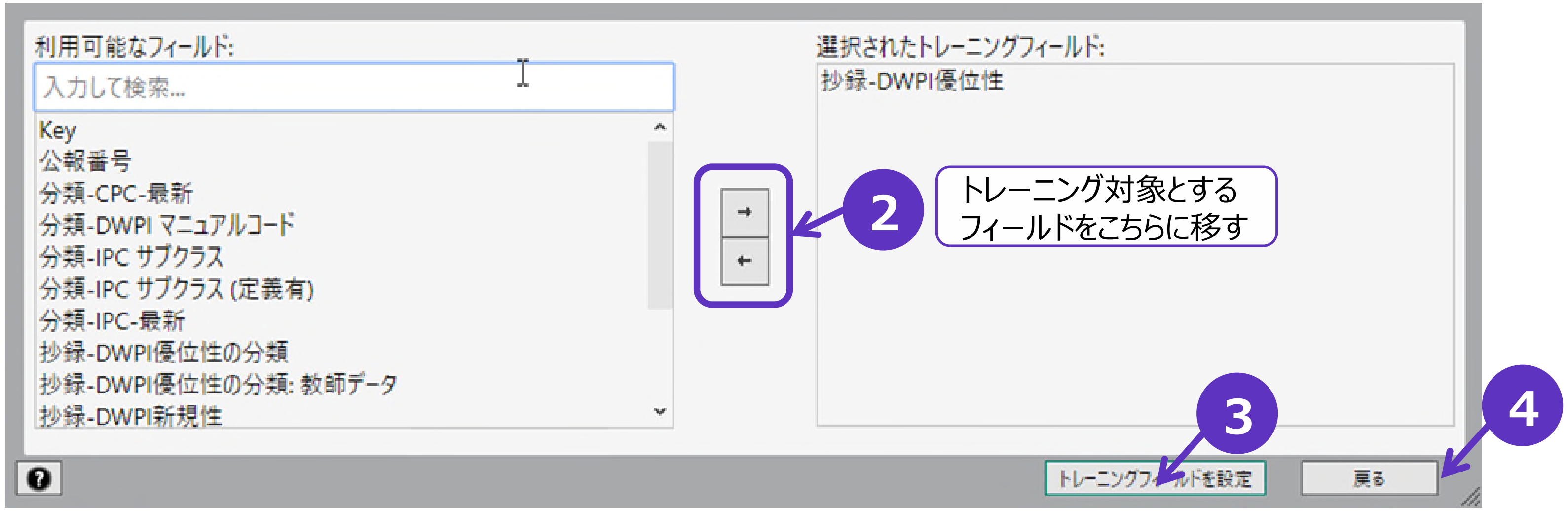
除外カテゴリを作成
「除外カテゴリ」は教師データの中に学習させたくないものがある場合に、それらを無視するものとして印を付けるカテゴリです。Disregardと表示されます。

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

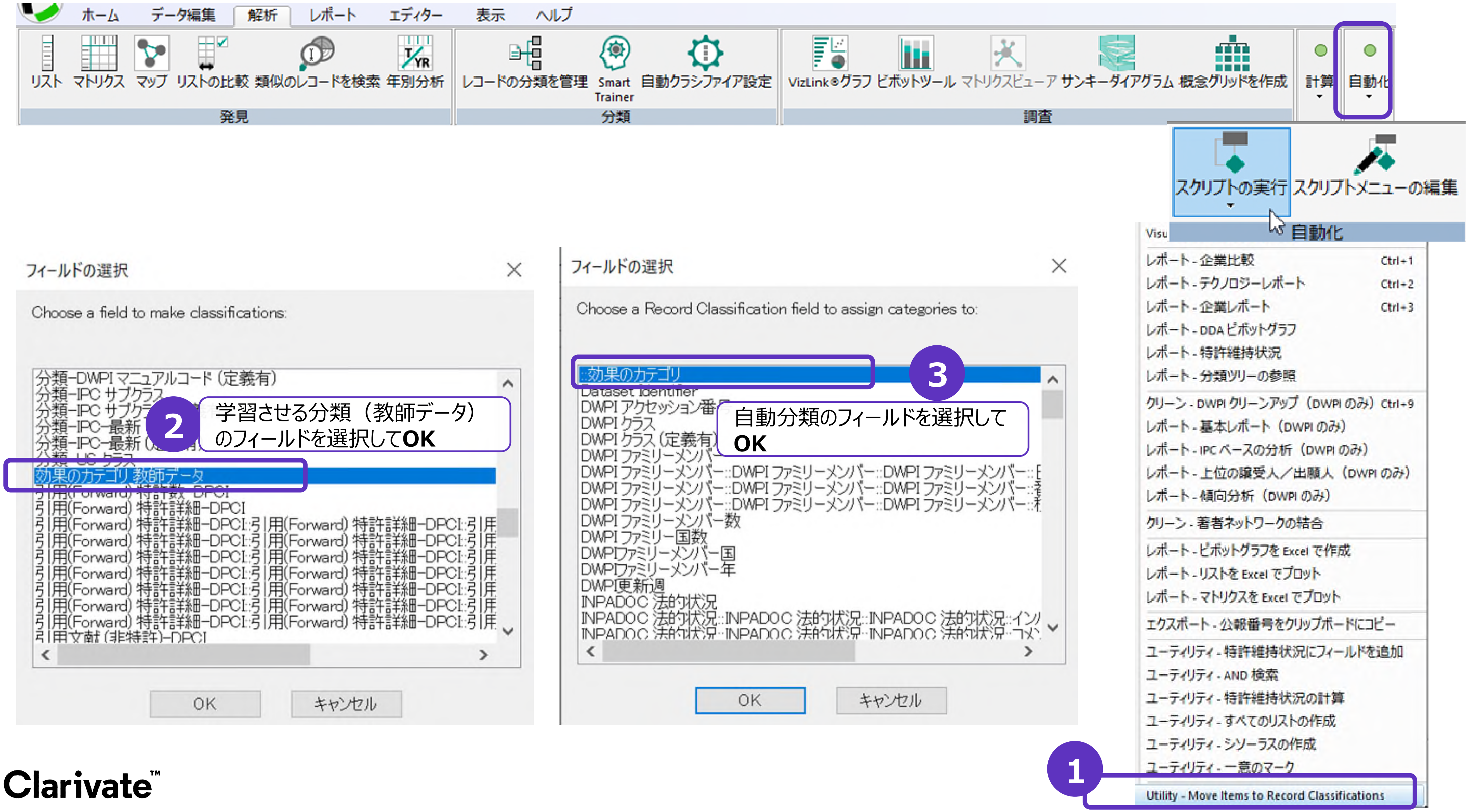






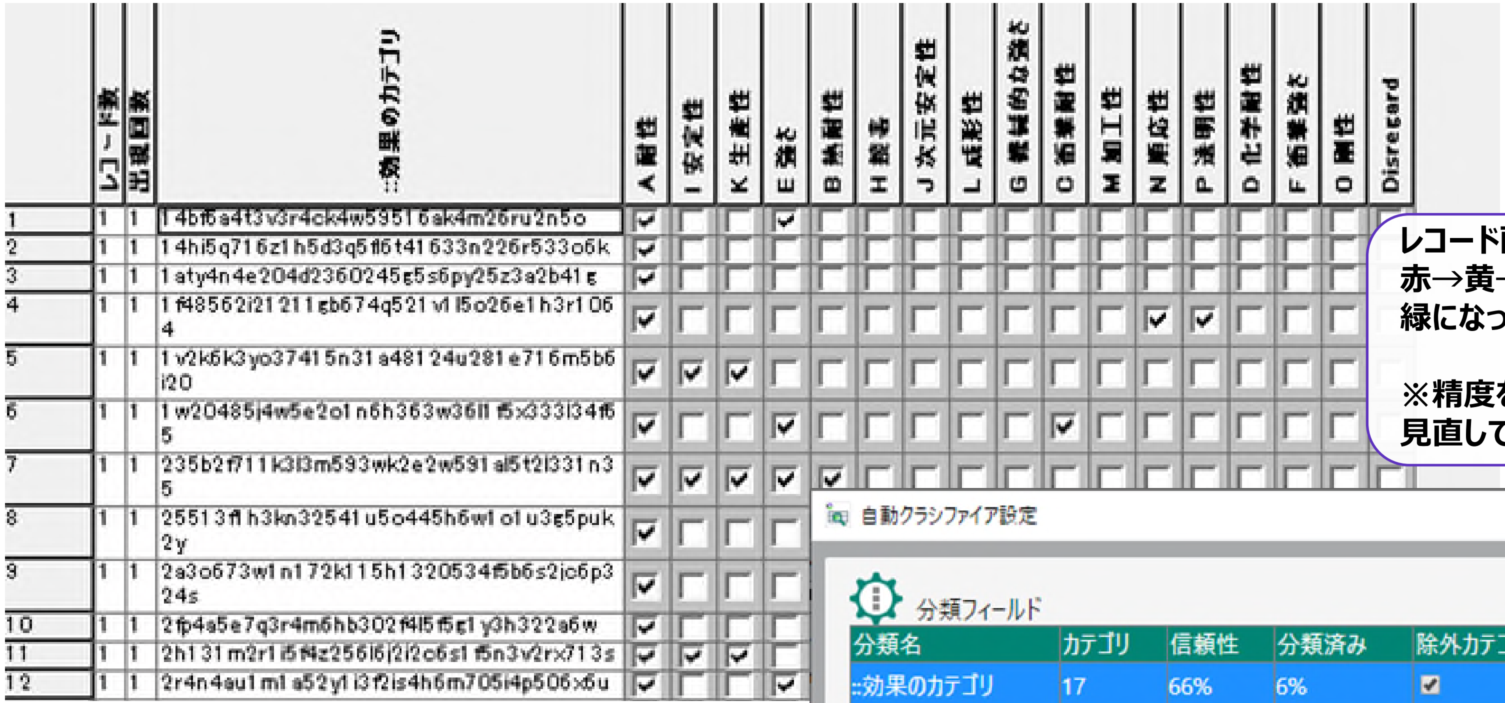
既存の分類データを利用 (教師データの作成はDDA外で行う)して、自動分類を行う

5． 教師データを読み込んで機械学習を行う



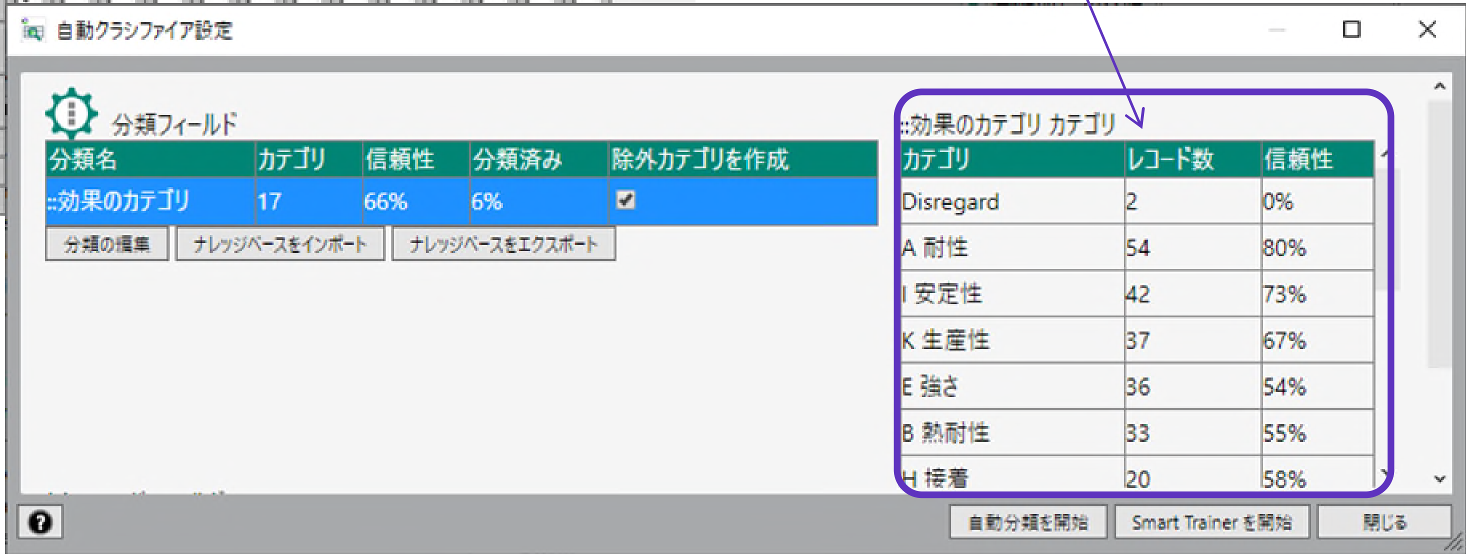
既存の分類データを利用 (教師データの作成はDDA外で行う)して、自動分類を行う

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。

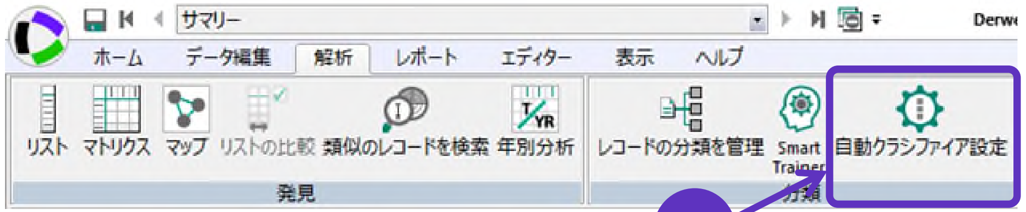


Clarivate™

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

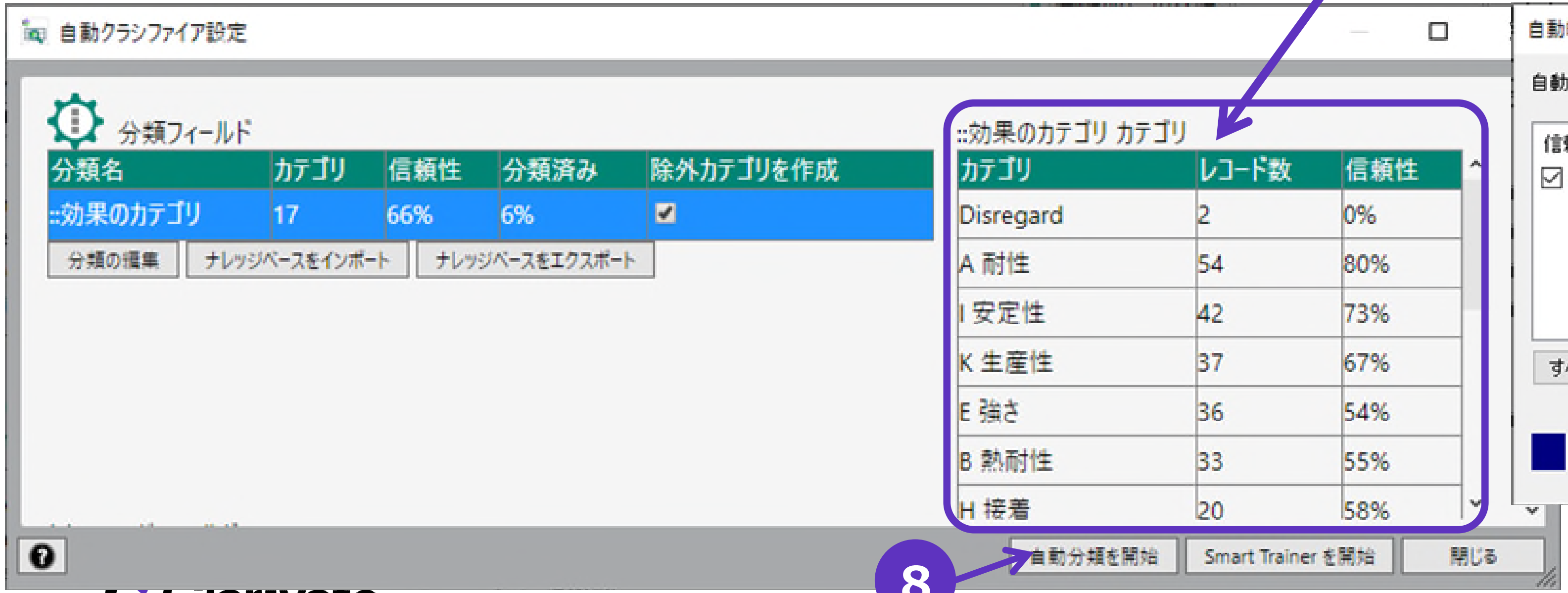
6．未分類データへ自動で分類を付与する

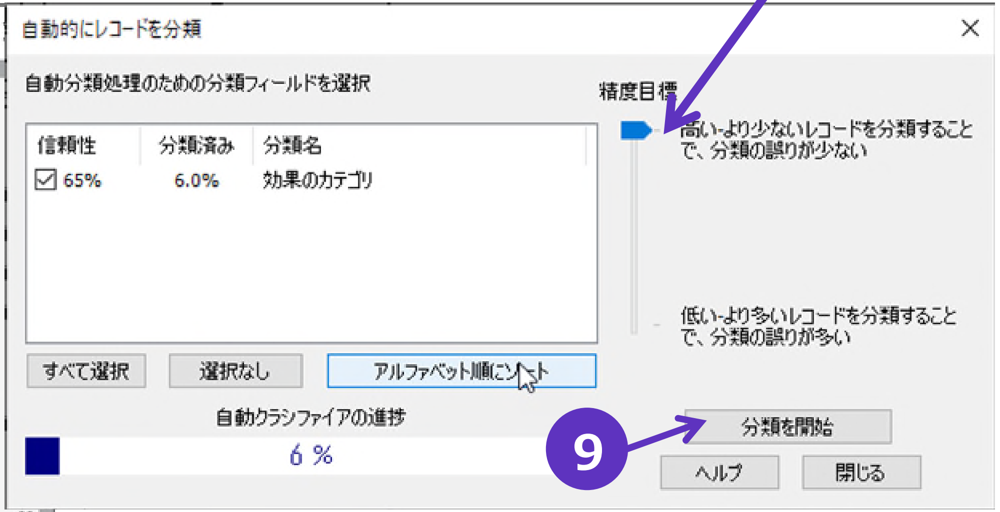
機械学習した内容で、自動で未分類のレコードを分類にカテゴライズします。
多くの手間がかかっていた分類付与を機械学習によって大幅に短縮することが可能です。



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。

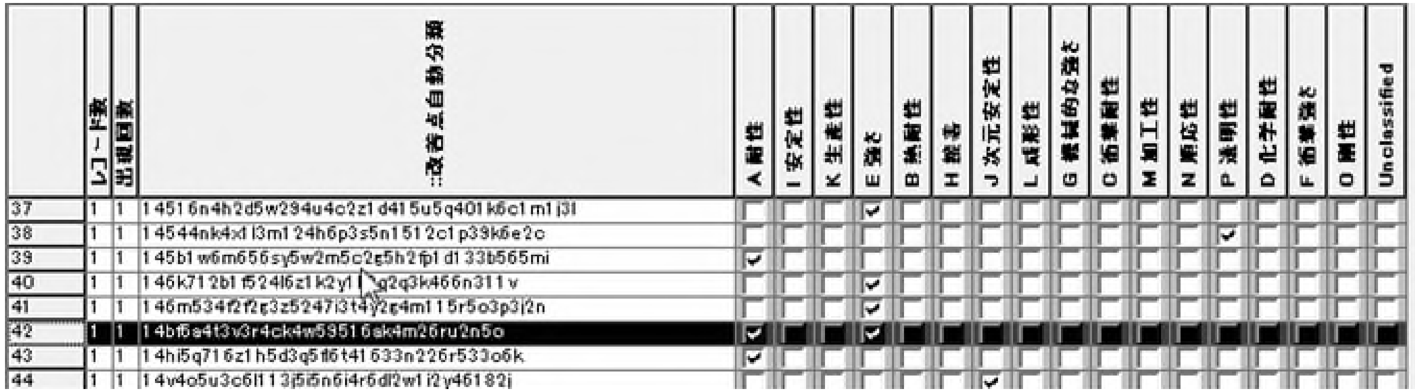
精度目標
精度低めでよりたくさんのレコードをカテゴリに組み込むか、精度を高くし、Uncategorizedに入るものが増えてもよいかの選択です。
精度高めにすることをお勧めします。

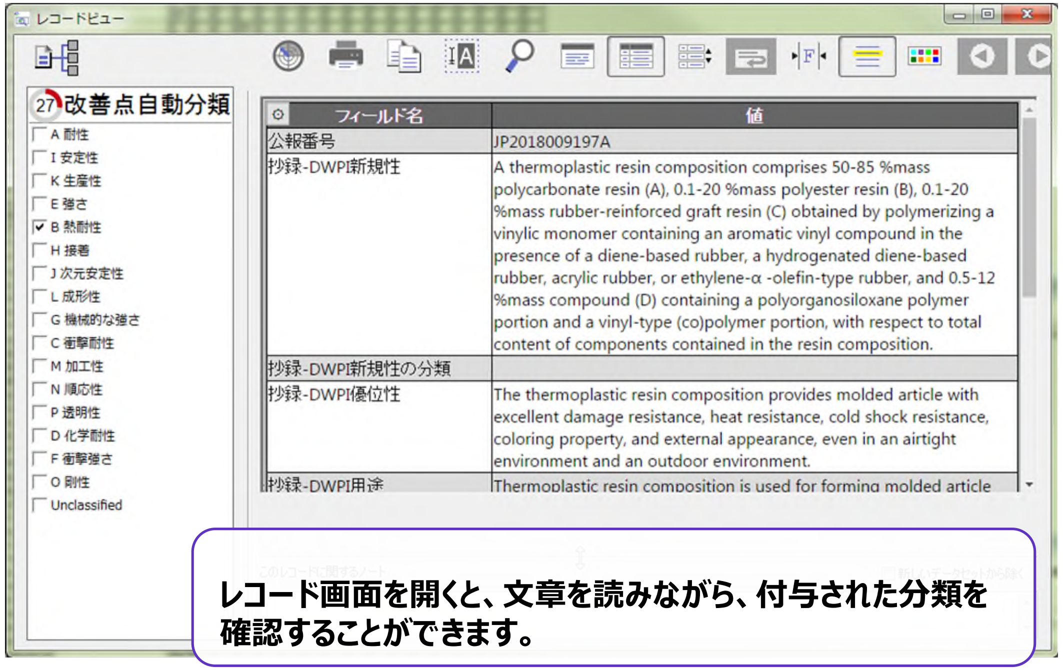




既存の分類データを利用 (教師データの作成はDDA外で行う)して、自動分類を行う

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。

Clarivate™

自動分類結果を用いた再分類

Clarivate™

自動分類結果による再分類：概要

自動分類前

リセット		抄録-DWPI 優位	1	2	3	4	5	6	7	8	9	10
		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	Show Values >= 1 and <= 32 共起 レコード数	2009	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

自動分類後(１回目)

リセット		::優位性	1	2	3	4	5	6	7	8	9	10	11
				1	5	58	68	68	12	16	98	9	4
優先権主張年（最先	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

606件のうち、506件を分類
(Unclassified除く)
※分類精度目標：高

自動分類後(2回目)

リセット		::優位性(インポ	1	2	3	4	5	6	7	8	9	10	11
			1	1	5	58	68	68	12	16	98	9	4
優先権主張年（最先	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		1

606件のうち、561件を分類
(Unclassified除く)
※分類精度目標：高

自動分類の結果をもとに再分類を行うことで、
Unclassifiedを削減

Clarivate™

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

↓

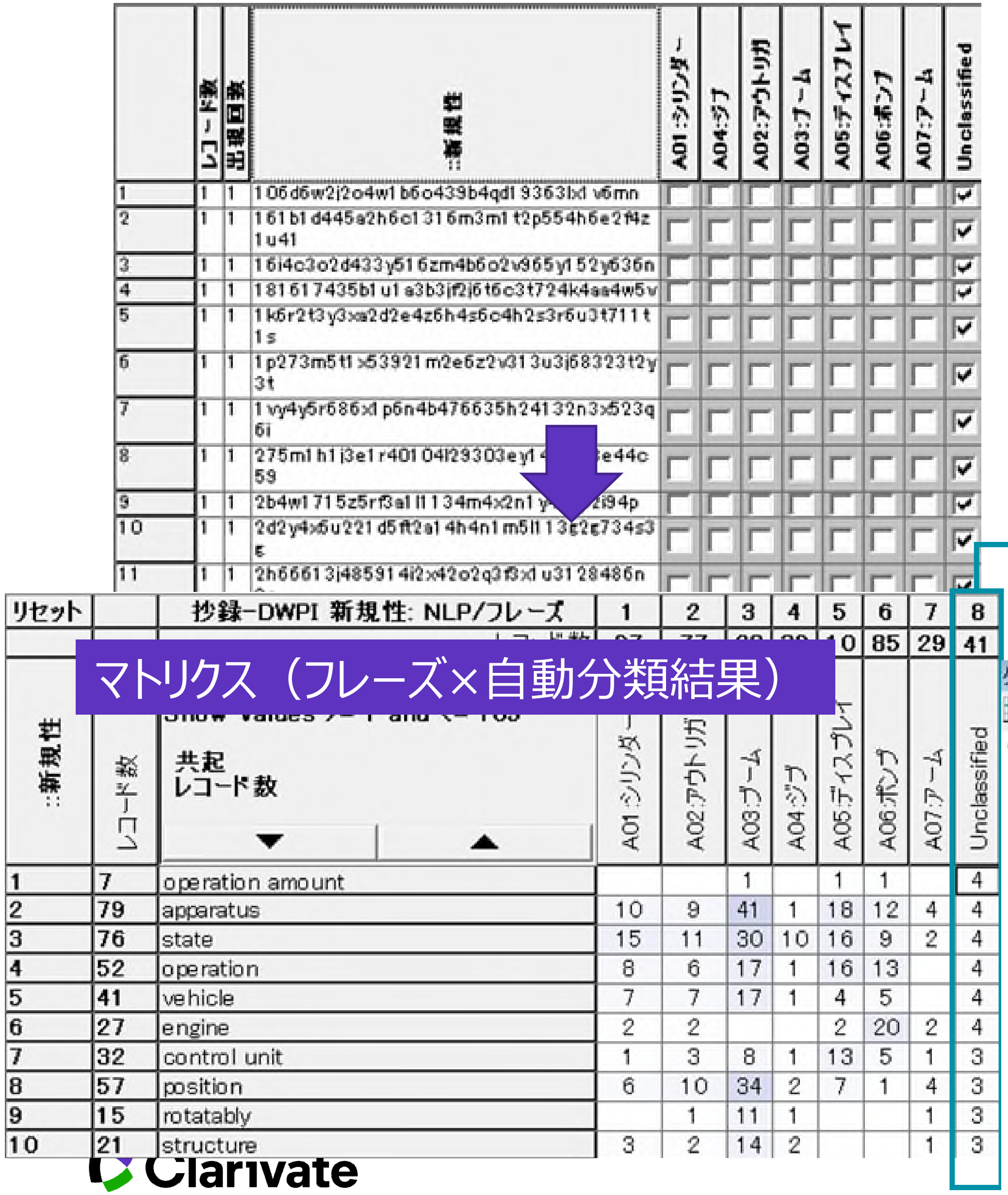
５．手順４でインポートして作成した分類カテゴリで再分類を行う

↓

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

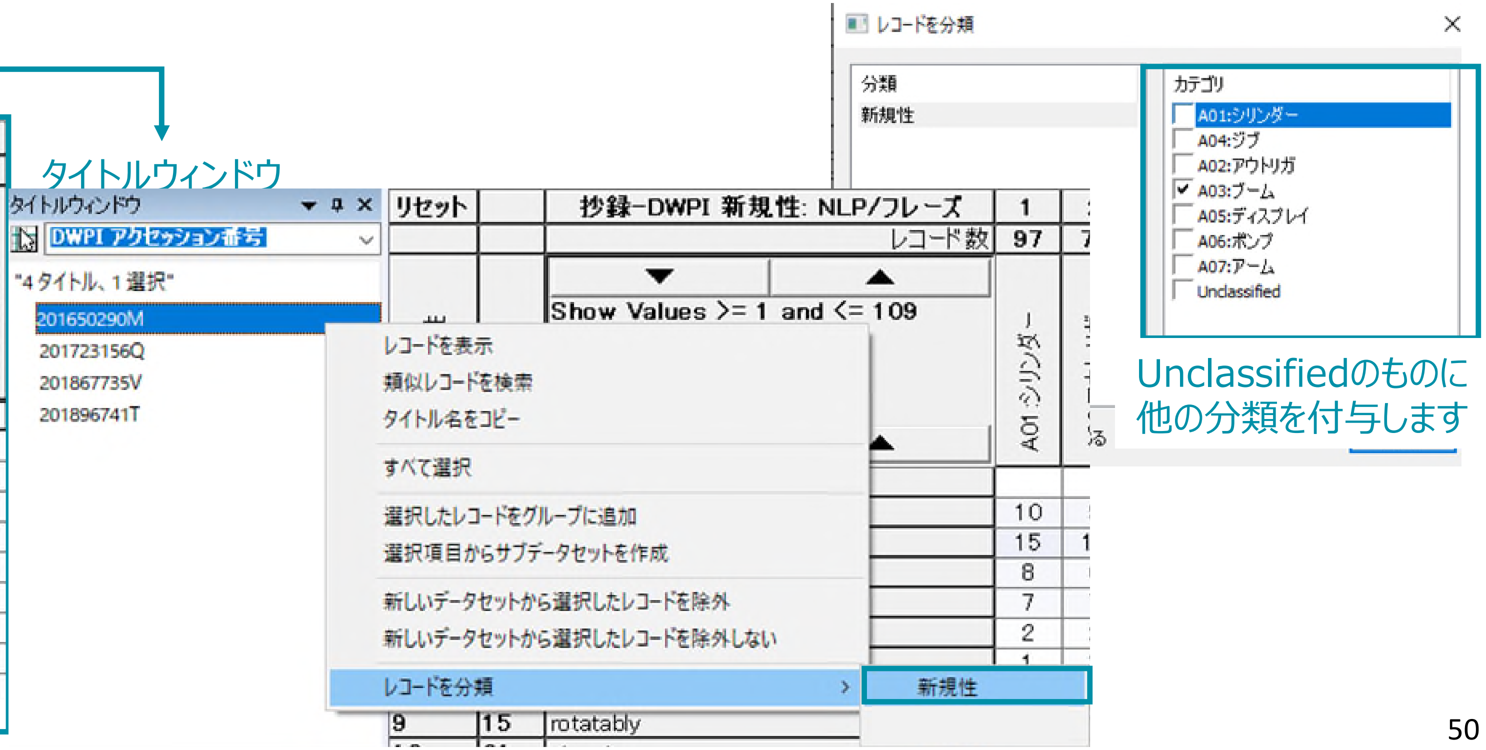
※手順１は必須ではありません。より正確な分類やUnclassifiedを多く削減する場合は実施すると有用です。



（手動分類付与方法①）

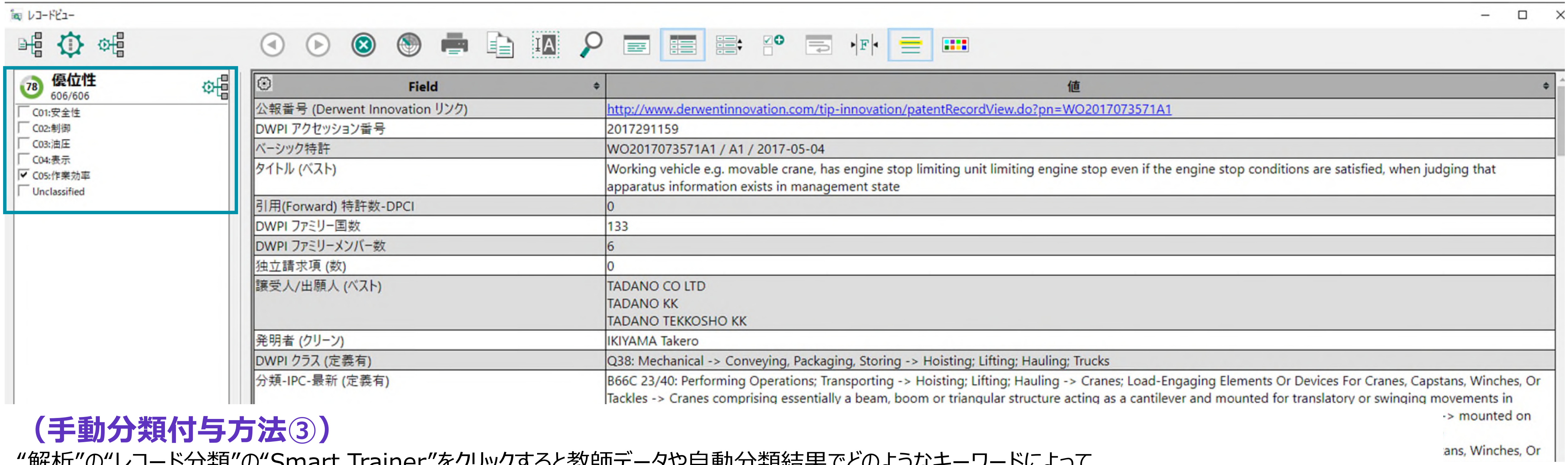
付与する際、フレーズ分解した文章（例では新規性:NLP/フレーズ）×自動分類した結果のマトリクスを作成し、

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。



（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。

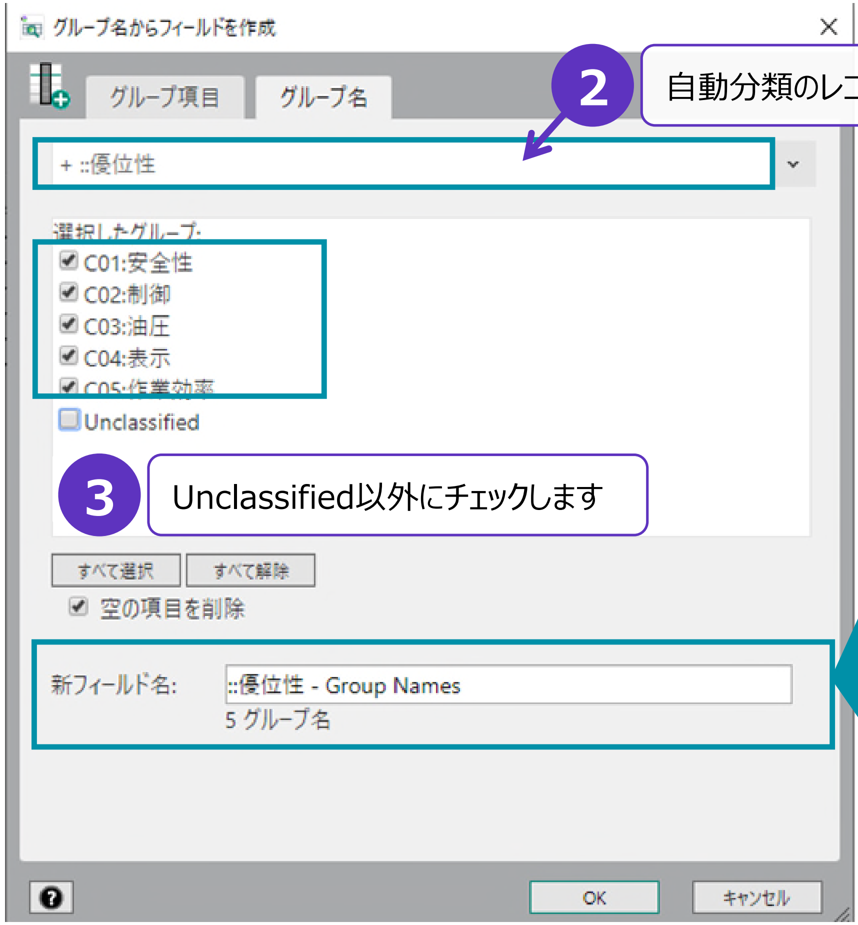


（手動分類付与方法③）

“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。

2．自動分類結果グループをフィールド化

1



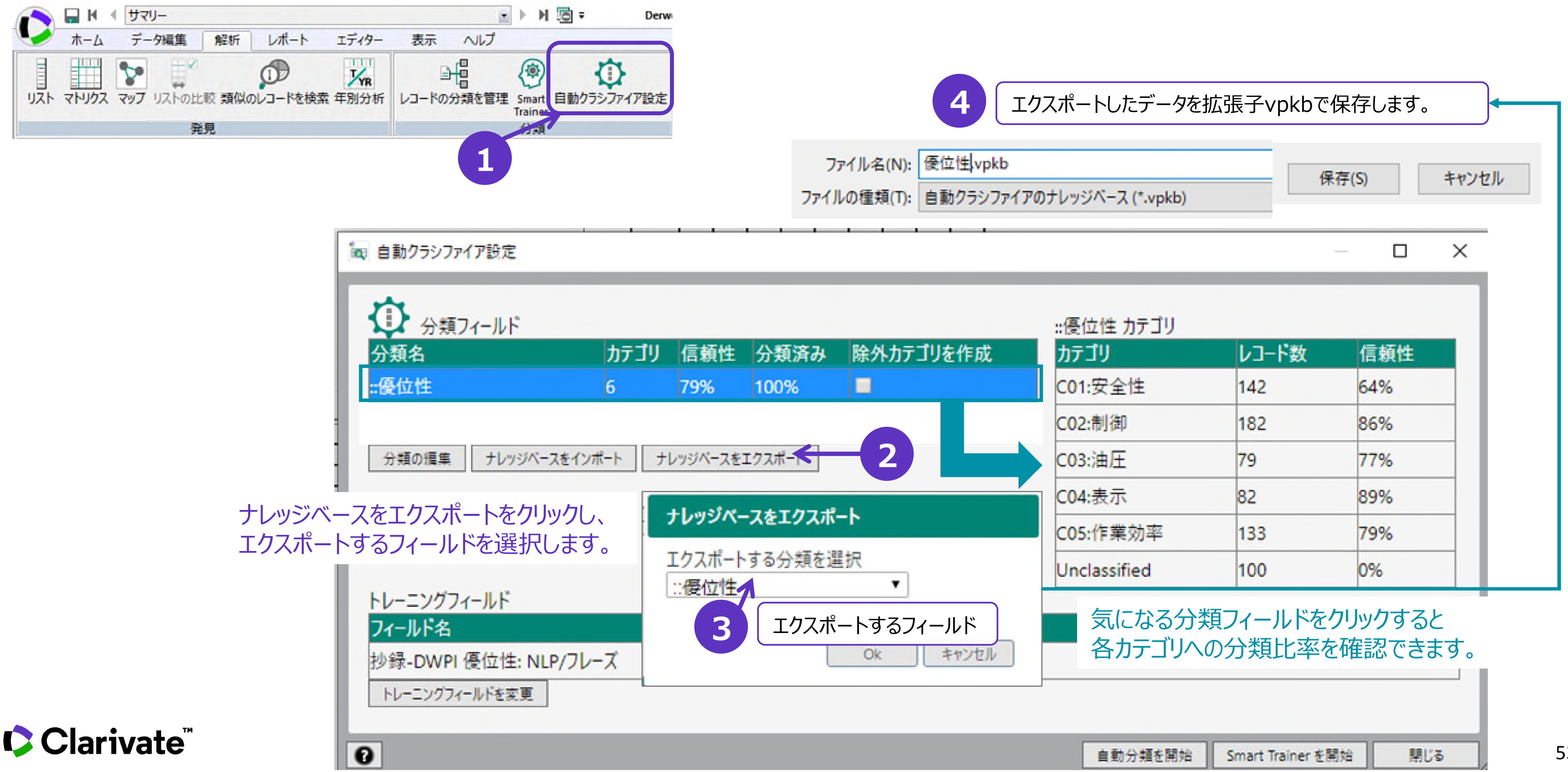
2 自動分類のレコード分類フィールドを選択します

3 Unclassified以外にチェックします

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

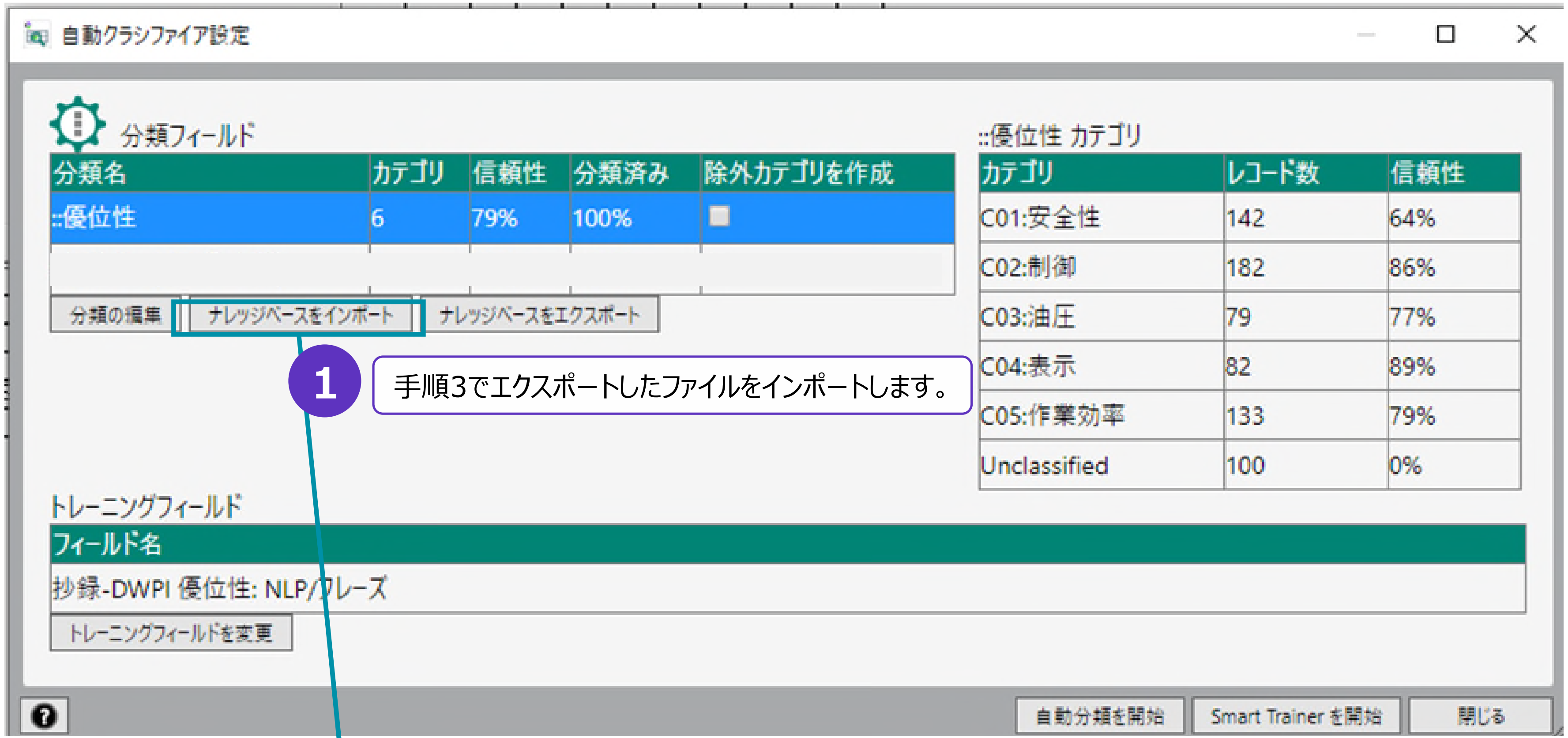
	レコード数	出現回数	::優位性 - Group Names
1	142	142	C01:安全性
2	182	182	C02:制御
3	79	79	C03:油圧
4	82	82	C04:表示
5	133	133	C05:作業効率

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



Clarivate™

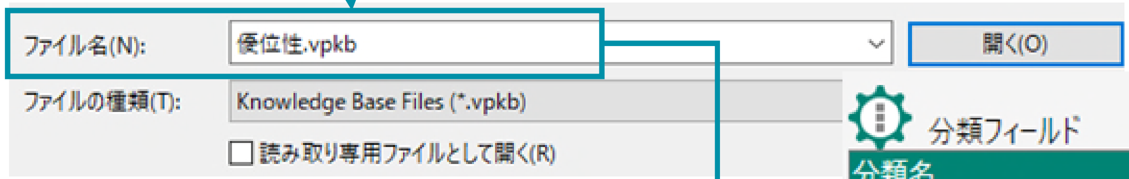
4．手順３でエクスポートしたファイルをインポートし、分類カテゴリを作成

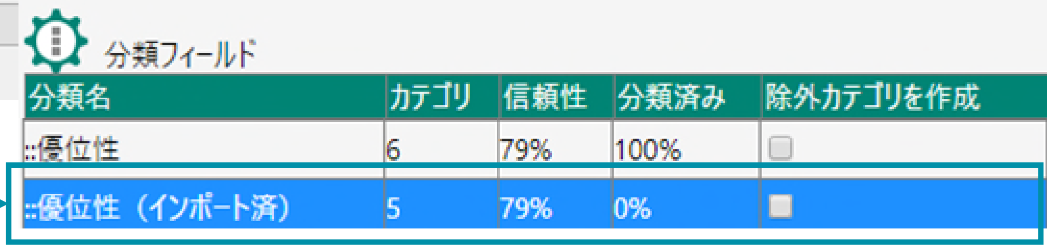


分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::優位性	6	79%	100%	☐

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%

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。

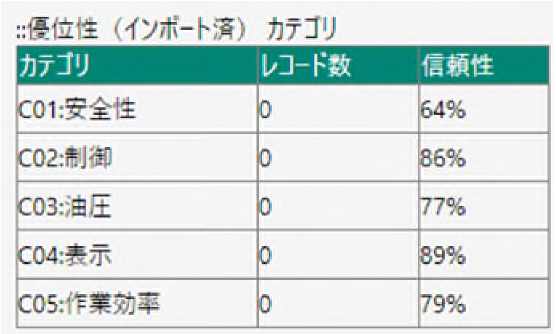




分類名	カテゴリ	信頼性	分類済み	除外カテゴリを作成
::優位性	6	79%	100%	☐
::優位性（インポート済）	5	79%	0%	☐

2 インポートが完了するとインポートした分類フィールドが作成されます。

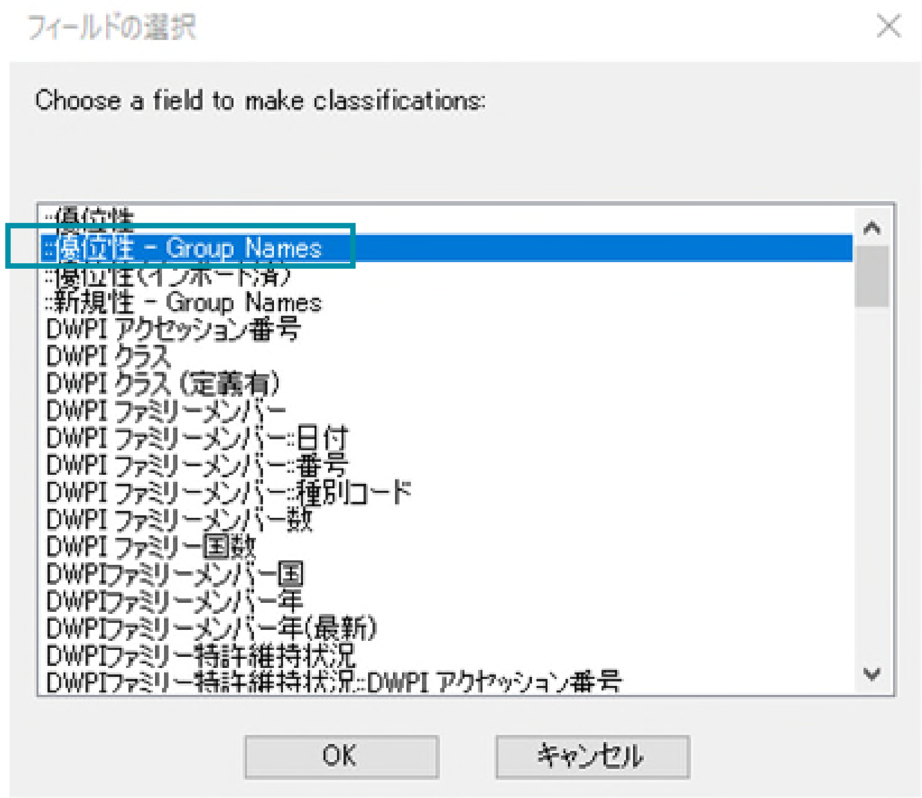
インポートした分類フィールドは分類がされていないので分類0%、レコード数0となっています。



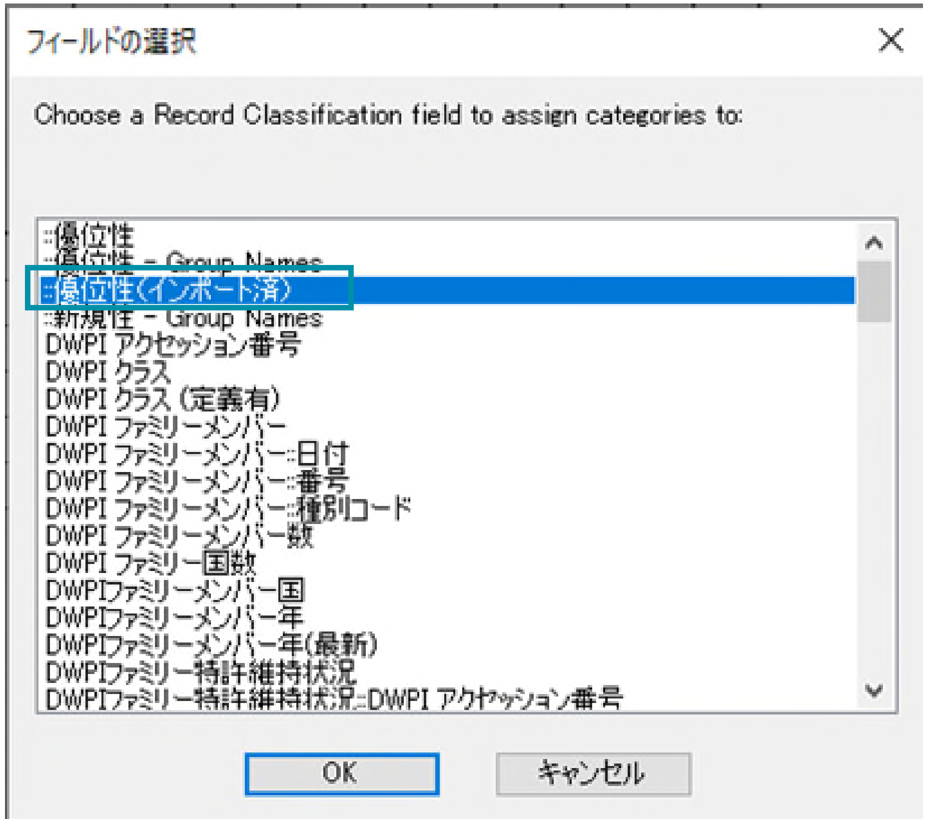
カテゴリ	レコード数	信頼性
C01:安全性	0	64%
C02:制御	0	86%
C03:油圧	0	77%
C04:表示	0	89%
C05:作業効率	0	79%

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

1

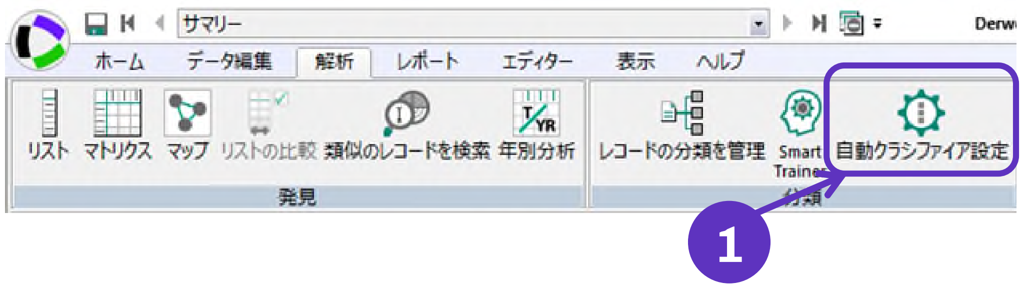


2 手順２で作成したフィールドを選択します。



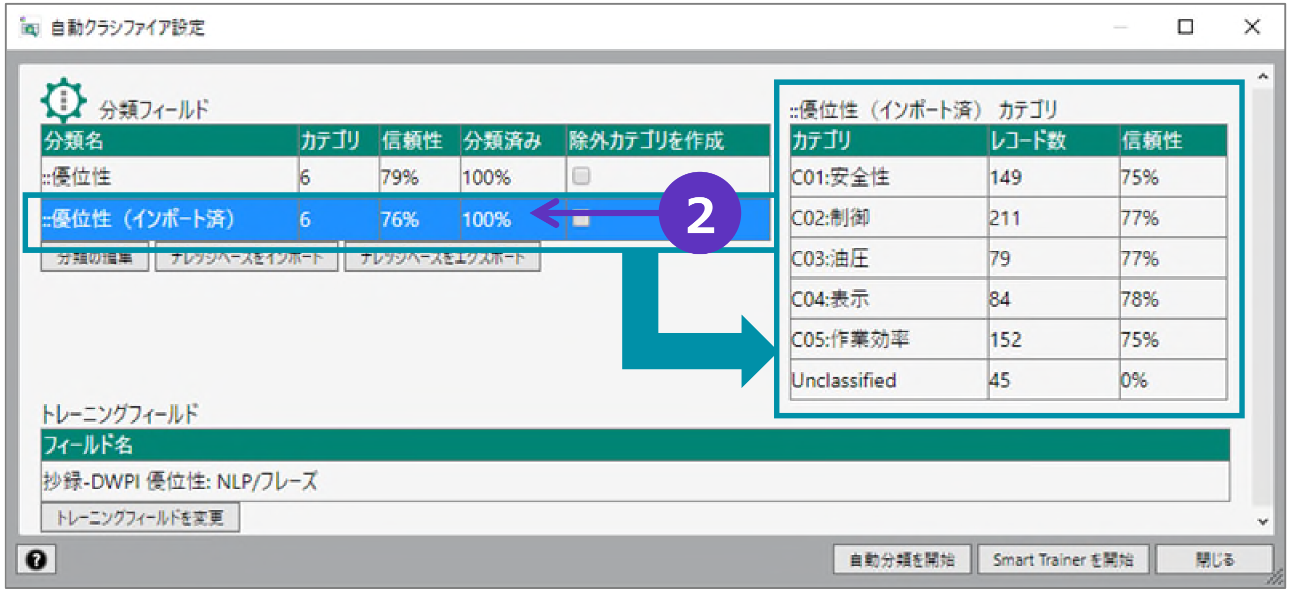
3 手順４でインポートした(インポート済)フィールドを選択します。

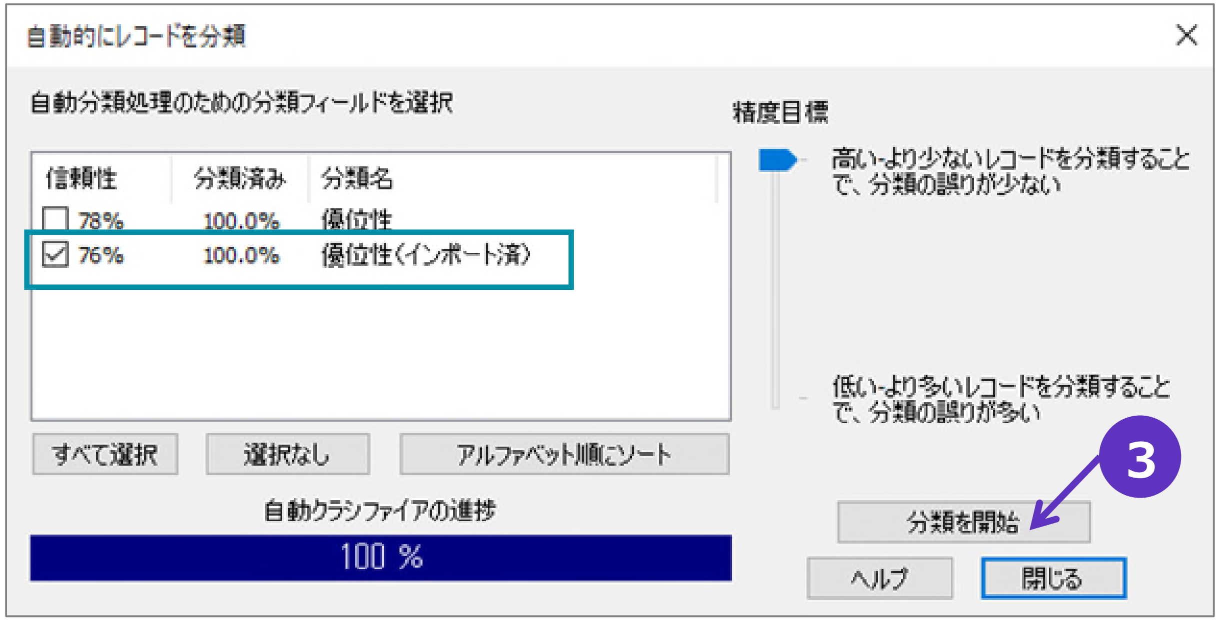
6．再度の自動分類を実施



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ァイ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。





Clarivate™

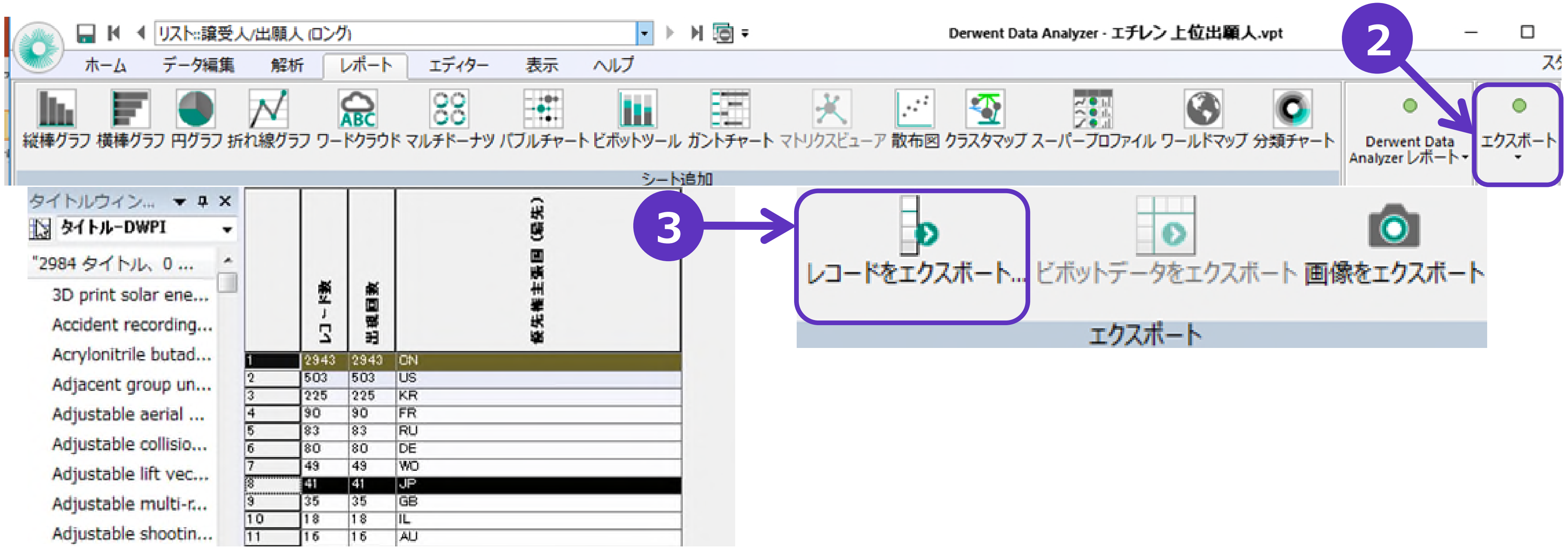
様々な処理：

・Excelへのデータ出力
・シソーラスの活用

Clarivate™

Excelへデータ出力

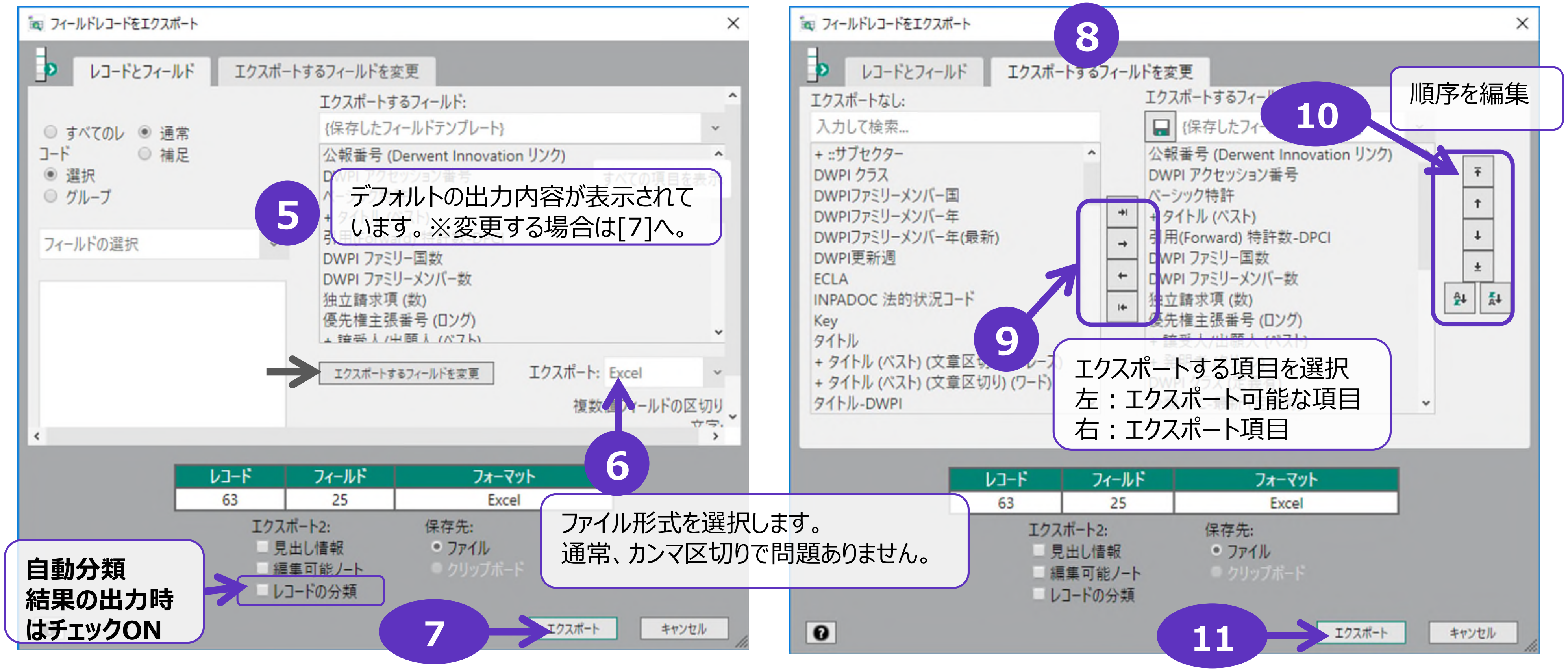
DDA上で様々データ整理・編集、グループ分け等を行ったデータは、他でも利用できるよう、様々な形式で出力することができます。



1

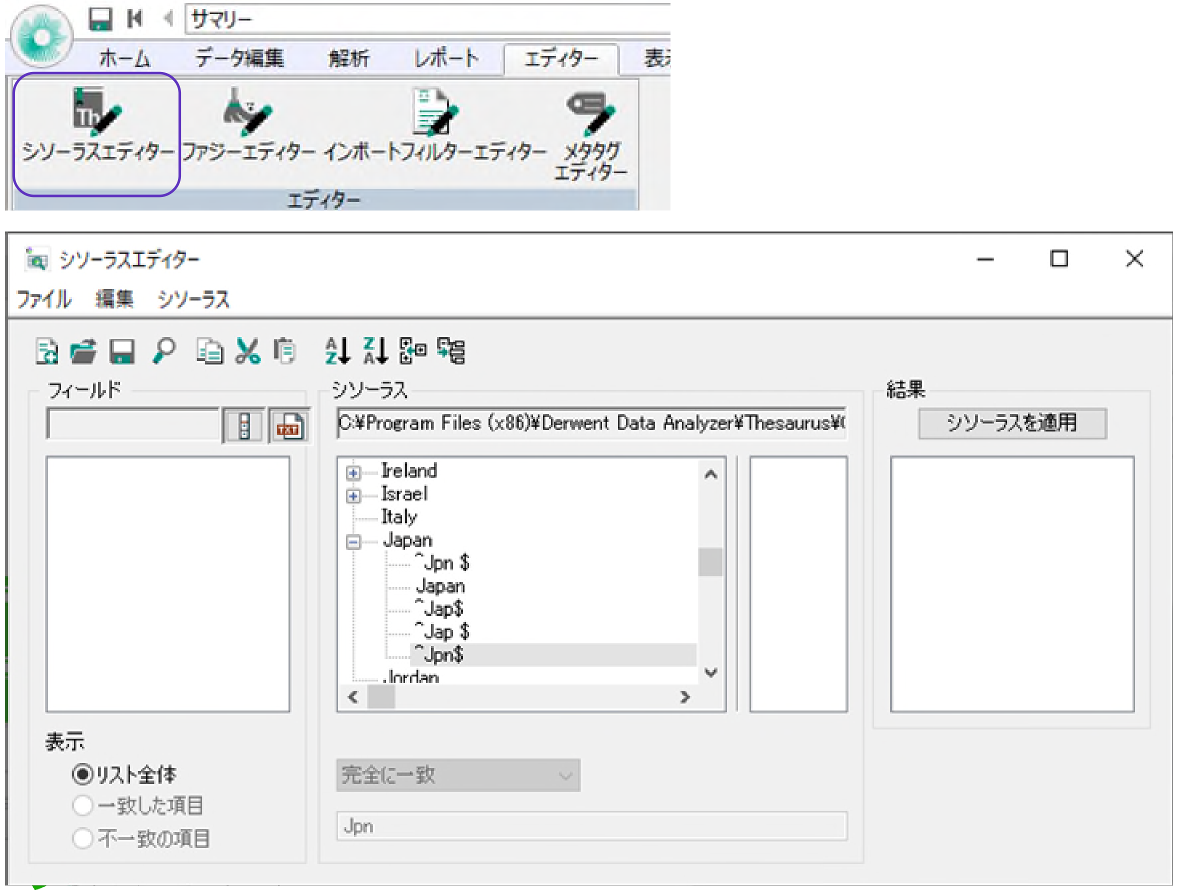
出力したいデータの範囲がある場合、
範囲を選択。
全レコードの場合は、必要ありません。

Excelへデータ出力

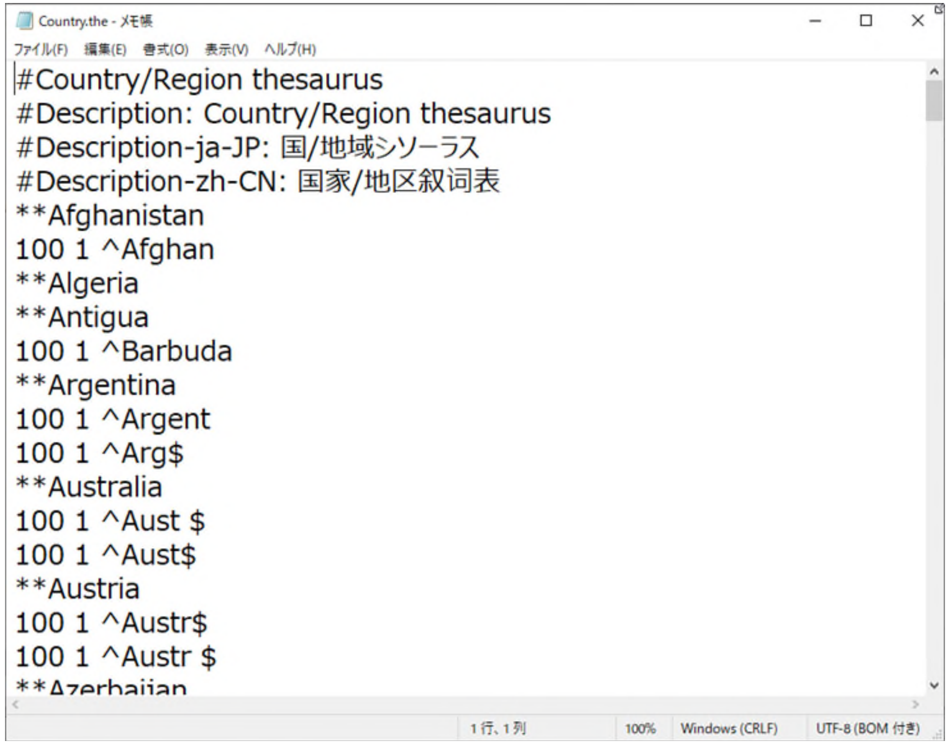


シソーラスの活用：シソーラスの独自編集

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
10	2	2	Automatic focus adjustment in microscope computes focussing ra
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
```

Clarivate™

Clarivate™

サービス全般に関するお問い合わせ

Tel:03-4589-3101

Think forward™

カスタマーケア（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：月ー金（祝祭日を除く）
午前9時30分～午後5時30分

About Clarivate

Clarivate™ is a leading global provider of transformative intelligence. We offer enriched data, insights & analytics, workflow solutions and expert services in the areas of Academia & Government, Intellectual Property and Life Sciences & Healthcare. For more information, please visit clarivate.com.

© 2024 Clarivate. All rights reserved
Clarivate and its logo, as well as all other trademarks used herein are trademarks of their respective owners and used under license.