



Cleaning Names with Vantage Point

Paul Oldham

Search Technology, Inc. thevantagepoint.com



Turn Information into knowledge

Home Products Support Downloads Resources Contact Us Login

search...



VantagePoint

Serious Software for Serious Professionals

VantagePoint is a powerful text-mining tool for discovering knowledge in search results from patent and literature databases. VantagePoint helps you rapidly understand and navigate through large search results, giving you a better perspective—a better vantage point—on your information. The perspective provided by VantagePoint enables you to quickly find WHO, WHAT, WHEN and WHERE, helping you clarify relationships and find critical patterns.

[Read more](#)

Product Presentations, Reviews, Case Studies, and Free Trials

[Introduction to VantagePoint](#) - Take a look at this presentation for a quick overview.

[VantagePoint Literature](#) - Download our 4-page brochure (also in **ESPAÑOL**, **PORTUGUÉS**, and **DEUTSCH**)

VantagePoint plays a role in reports from the UK Intellectual Patent Office.

"Will VantagePoint work with my data?"

VantagePoint comes with access to our library of import filters, thesauri, and macros. These provide the resources you need to import, transform, analyze, and visualize data from a large number of on-line providers. - [Browse the Library](#) - VantagePoint comes with tools to build your own import filters, thesauri, and macros for your proprietary data sources and business processes.

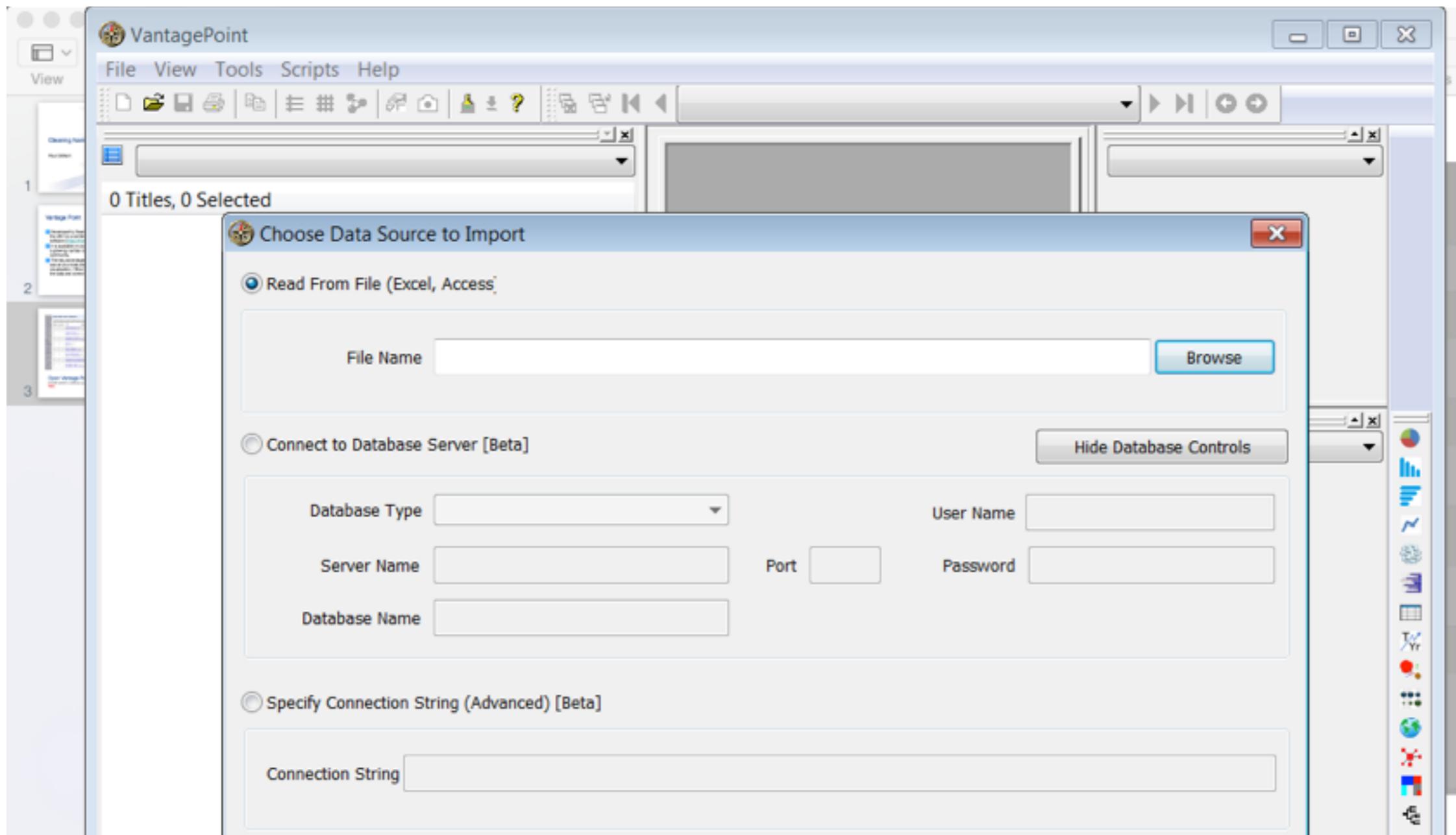
News

[Upcoming OECD BLUE SKY III Presentation](#)

[Just Published: How Multidisciplinary Are the Multidisciplinary Journals Science and Nature?](#)

Open Vantage Point and Register

- Enter the registration code
- Enter your email address
- This is a 1 week code and will expire



Open Vantage Point

Choose Cancel to close pop up. Then **File > Import Database Table. Browse to drones_tac_1606_families.**

The screenshot shows the WIPO Lens software interface. The title bar reads "Summary". The left sidebar shows a tree view with "Title" selected, and a message "0 Titles, 0 Selected". The main area displays a "SUMMARY SHEET" for a dataset with 1,606 records. The top right of the summary area shows "Number of Records: 1,606" and buttons for "Columns" and "I". Below this is a table with columns: "Field", "Number of Items", "% Coverage", "Data Type", and "Meta Tags". The table includes a filter row "(filters) >=500" and a header row. Data rows include:

Field	Number of Items	% Coverage	Data Type	Meta Tags
(filters)	>=500			
#	1,606	100%		
Applicants	1,154	98%		
Application Date	1,265	99%		
Application Number	1,606	100%		
Cited Count	49	100%		
CPC Classifications	914	60%		
DOI(s)	51	3%		
Extended Family Size	43	100%		
Full Text	2	100%		
Inventors	1,382	95%		
IPC Classifications	1,289	97%		

A vertical toolbar on the right contains icons for various data analysis and visualization tools.

The Summary Sheet

Double click on a field to open it. Try double click on **Applicants**

VantagePoint - [VantagePoint1]

File Edit View Sheets Fields Groups Tools Scripts Window Help

List::Applicants

Title

29 Titles, 0 Selected

Accessoire Pour Rendre Un Drone A...
Altitude Estimator For Rotary-wing ...
Appareil Mobile, Notamment Dron...
Balise De Jalonnement Pour L'orient...
Carrier Block For Motor Of Rotary-...
Drone A Voiture Tournante Avec He...
Drone A Voiture Tournante De Type...
Drone A Voiture Tournante Muni D'...
Einrichtung Zum Fliegen Einer Drone
Electronic Navigation Card Holder ...
Equipement De Telecommande De ...
Intuitive Piloting Method Of A Dron...
Method For Detection Of External ...
Method For Piloting A Rotary-wing ...
Method For Piloting A Rotary-wing ...
Method For Piloting A Rotary-wing ...
Method For Synchronised Control O...

Records # Instances Applicants

	# Records	# Instances	Applicants
1	29	29	PARROT
2	27	27	THALES SA
3	14	14	IDEAL TOY CORP
4	13	13	DCNS
5	10	10	MATSUSHITA ELECTRIC IND CO LTD
6	9	9	US NAVY
7	8	8	BOEING CO
8	8	8	G MASHINOSTROITEL NOE KB RADUGA IM A JA
9	7	7	ALSHDAIFAT WASFI;; ALMUHAIRBI EIDA;; KASSA
10	7	7	BAODING LONG TENG SPORTS GOODS CO LTD
11	7	7	EUROCOPTER FRANCE
12	7	7	FLIR SYSTEMS
13	7	7	HYUNDAI MOTOR CO LTD
14	7	7	OBSHESTVO S OGRANICHENNOJ OTVETSTVEN
15	6	6	BOSE CORP
16	6	6	DELTA DRONE
17	6	6	DIEHL GMBH & CO
18	6	6	DUROV DMITRIJ SERGEEVICH

Publication Year

	Publication Year
7	2016
6	2012
6	2015
5	2013
3	2011
1	2010
1	2014

Applicants View

For each applicants we see the title. We can also add a field in the Detail Window (right) by clicking on it. Add new detail windows and move around to suit you (View > Add Detail Window). This is the power of Vantage Point... we can see a lot at once.

PREPARING TO CLEAN DATA BY SPLITTING FIELDS

Source File: E:\WIPO Training\training_datasets\lens_drones_tac\dron...
Source Date: Sep 16 2016 12:23
Source Database: Quick Import: drones_tac_1606_families - drones_t...

SUMMARY SHEET

Number of Records: 1,606

Field

(filters)

#

Applicants

- Applicants
- Create List
- List Cleanup...
- Thesaurus...
- Find and Replace...

Further Processing >

- :Read All
- Apply Proper Case
- Author Cleanup
- Dates
- Divide at
- Extract Authors
- Extract Country
- NLP
- Patent Numbers
- Read After - First
- Read After - Last
- Read Until - First
- Read Until - Last
- Remove

Colon

Comma

New Line

Period

Pipe (|)

Punctuation Mark

Semicolon

Underscore

Whitespace

Summary List:Applicants

VantageP...

Tools Scripts Window Help

Summary

Source File: E:\WIPO Training\training_datasets\lens_drones_tac\dron...
Source Date: Sep 16 2016 12:23
Source Database: Quick Import: drones_tac_1606_families - drones_t...

SUMMARY SHEET

Number of Records: 1,606

Field

(filters)

#

Applicants

Applicants: Divide at/Semicolon

Application Date

Application Number

Cited Count

CPC Classifications

DOI(s)

Summary List:Applicants List:Applicants: Divide at/Semicolon

VantageP...

Right click applicants. **Further processing > Divide at > Semicolon.** Creates new field in summary

Repeat for other fields (Right click in summary)

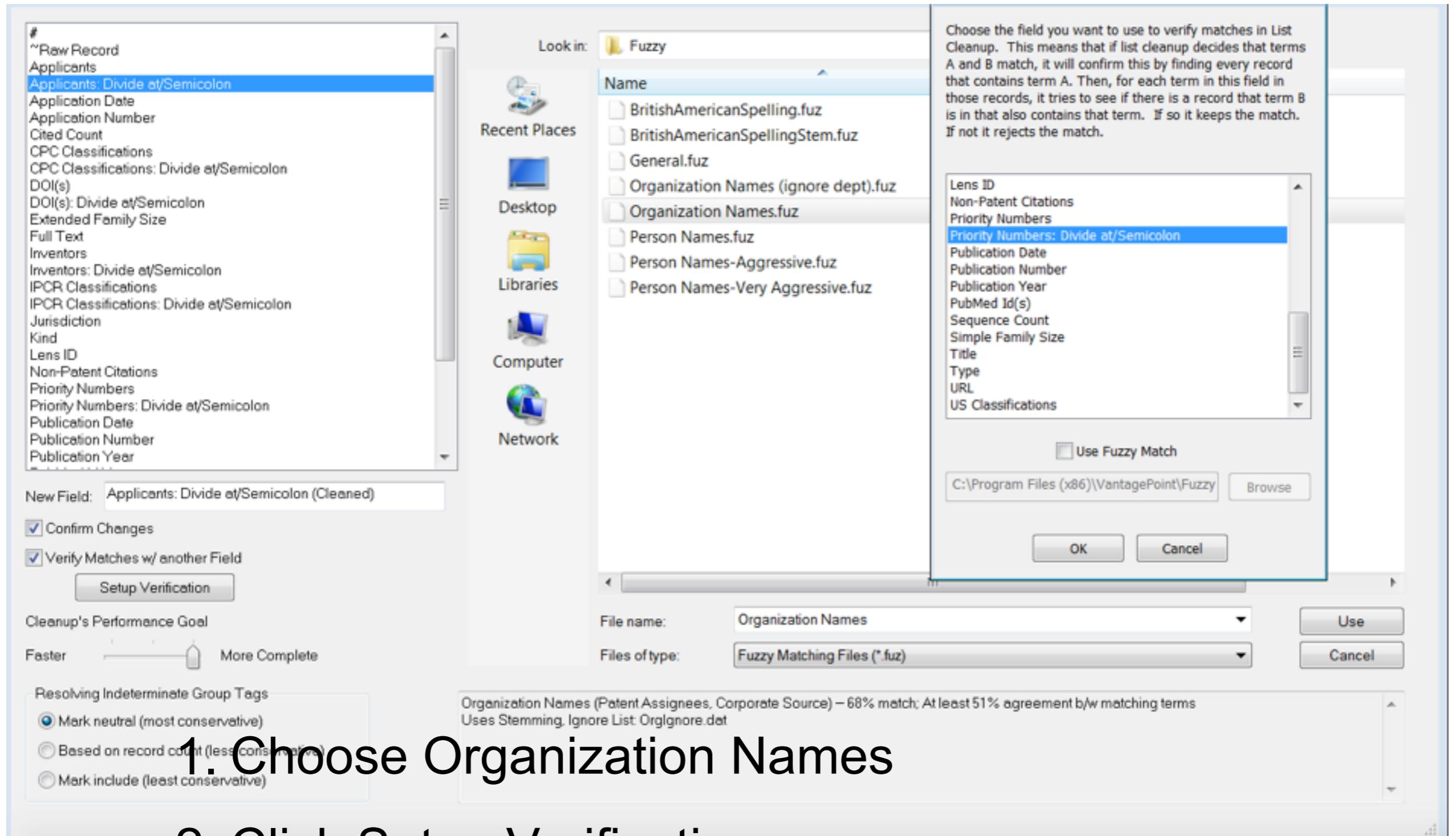
- **Dois** > Further Processing > Divide At > Semicolon
- **Inventors** > Further Processing > Divide At > Semicolon
- **IPC Classifications** > Further Processing > Divide At > Semicolon
- **Priority Numbers** > Further Processing > Divide At > Semicolon
- Then
- File > Save > drones_tac_1606_families_yourinitials

The screenshot shows a software application window titled "Summary". The top menu bar includes File, Edit, View, Sheets, Fields, Groups, Tools, Scripts, Window, and Help. The left sidebar has a title filter set to "Title" with "0 Titles, 0 Selected". The main area displays a "SUMMARY SHEET" with the text "Number of Records: 1,606". A context menu is open over the "Applicants" field, listing options: Create List, List Cleanup..., Thesaurus..., Find and Replace..., Further Processing >, Extract My Keywords >, Rename Field..., and Copy Field... . The right sidebar shows a "Publication Year" filter. The bottom right corner features a vertical toolbar with various icons.

In Summary Right Click Applicants: Divide At Semicolon > List Cleanup

The importance of the divided fields will become apparent in a moment.

List Cleanup Panel



1. Choose Organization Names

2. Click Setup Verification

3. Scroll to Priority Numbers: Divide at semicolon

Cleanup Confirm

Cleanup Sets Detail Window

Item Name	Num Records
ZHOU ANDREW H B	5
ZHOU ANDREW	2
ZHOU ANDREW H B	3
ZHOU TIGER T G	5
ZHOU TIGER	2
ZHOU TIGER T G	3
THOMAS MICHAEL STEVEN	3
THOMAS MICHAEL S	1
THOMAS MICHAEL STEVEN	2
UBIQOMM LLC	3
REDDYS LAB LTD DR	2
SHENZHEN AEE AVIAT TECHNOLOGY CO LTD	2
SHENZHEN AEE AVIAT TECHNOLOGY CO LTD	2
SHENZHEN AEE INTELLIGENT TECHNOLOGY CO LTD	1
SHENZHEN AEE TECHNOLOGY CO LTD	2
SKYPERSONIC	2
MTN SATELLITE COMM	1
SCHLUMBERGER TECHNOLOGY BV	1
SIMMS JOHN	1
SIMMS JOHN	1
SIMMS JR JOHN	1

Display

All Items
 Combined Items
 Custom Set of Items

Custom Set of Items

Find Close Matches (50%)

Remove All Invert Set

Find

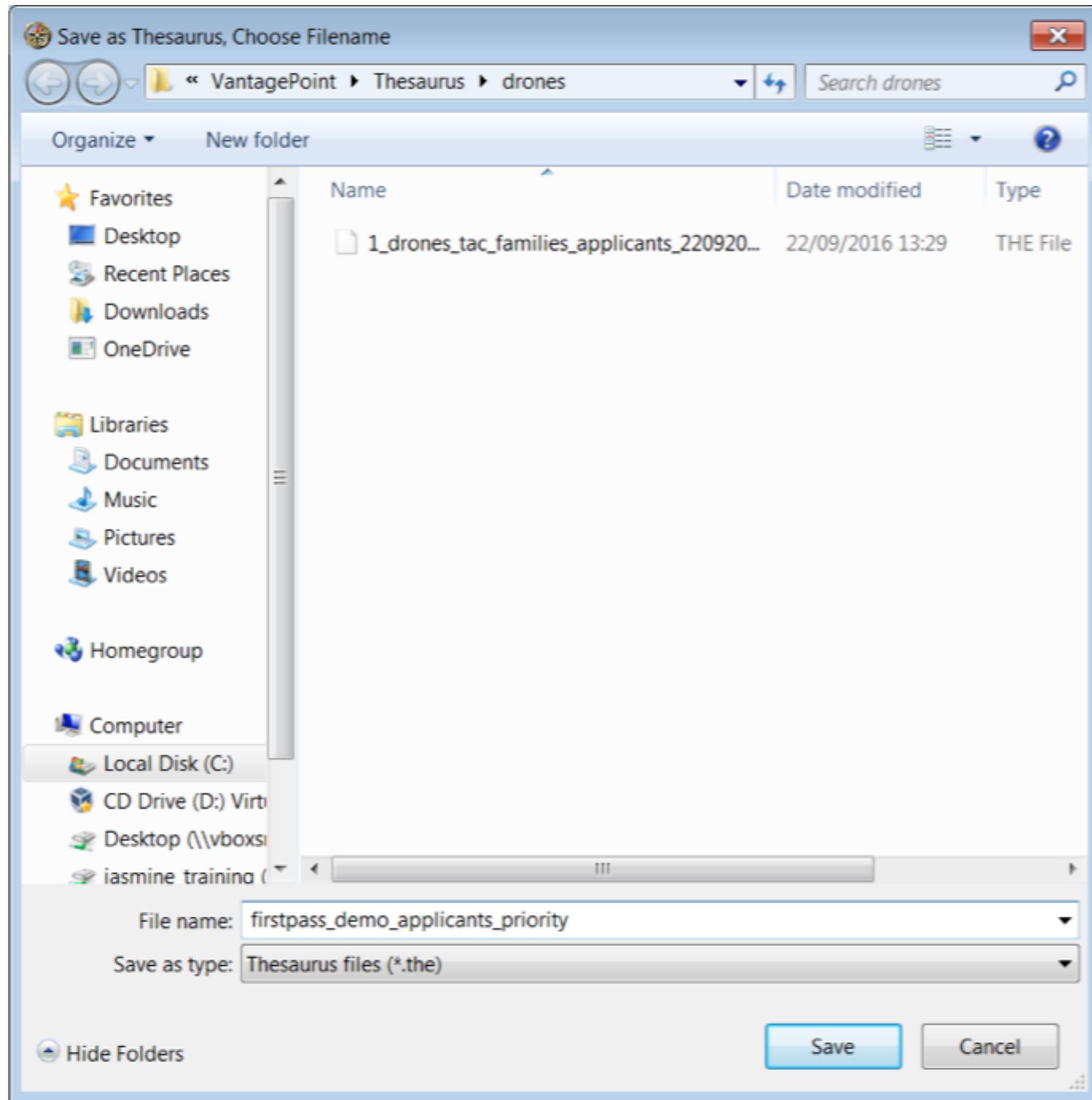
Regular Expression

Add Remove

Save as Thesaurus

Accept

Cancel



Cleanup Confirm

Cleanup Sets Detail Window

Item Name	Num Records
ZHOU ANDREW H B	5
ZHOU ANDREW	2
ZHOU ANDREW H B	3
ZHOU TIGER T G	5
ZHOU TIGER	2
ZHOU TIGER T G	3
THOMAS MICHAEL STEVEN	3
THOMAS MICHAEL S	1
THOMAS MICHAEL STEVEN	2
UBIQOMM LLC	3
UBIQOMM	1
UBIQOMM LLC	2
REDDYS LAB LTD DR	2
REDDYS LAB INC DR	1
REDDYS LAB LTD DR	2
SHENZHEN AEE AVIAT TECHNOLOGY C...	2
SHENZHEN AEE AVIAT TECHNOLOG...	2
SHENZHEN AEE INTELLIGENT TECH...	1
SHENZHEN AEE TECHNOLOGY CO L...	2
SKYPERSONIC	2
SKYPERSONIC	1
SKYPERSONIC LLC	1
MTN SATELLITE COMM	1
MTN SATELLITE COMM	1
MTN SATELLITE COMM INC	1
SCHLUMBERGER TECHNOLOGY BV	1
SCHLUMBERGER TECHNOLOGY BV	1
SCHLUMBERGER TECHNOLOGY CO...	1
SIMMS JOHN	1
SIMMS JOHN	1
SIMMS JR JOHN	1

Display

All Items
 Combined Items
 Custom Set of Items

Custom Set of Items

Find Close Matches (50%)

Remove All Invert Set

Find

Regular Expression

Add Remove

Save as Thesaurus

Accept

Cancel

SUMMARY SHEET

Number of Records: 1,606

Columns

i

Field	Number of Items	% Coverage	Data Type	Meta Tags
(filters)	>=500			
#	1,606	100%		
Applicants	1,154	98%		
Applicants: Divide at/Semicolon	1,396	98%		

[Applicants: Divide at/Semicolon \(Cleaned\)](#)

Application Date

Application Number

Cited Count

CPC Classifications

CPC Classifications: Divide at

DOI(s)

DOI(s): Divide at/Semicolon

Extended Family Size

Full Text

Inventors

Inventors: Divide at/Semicolon

IPC Classifications

Create List

List Cleanup...

Thesaurus...

Find and Replace...

Further Processing

Extract My Keywords

Rename Field...

Copy Field...

Set Data Type

Set Meta Tags...

Delete Field...

Use Applicants with
Cleaned At the end >

List Cleanup

List Cleanup

 ~Raw Record
 Applicants
 Applicants: Divide at/Semicolon
Applicants: Divide at/Semicolon (Cleaned)
 Application Date
 Application Number
 Cited Count
 CPC Classifications
 CPC Classifications: Divide at/Semicolon
 DOI(s)
 DOI(s): Divide at/Semicolon
 Extended Family Size
 Full Text
 Inventors
 Inventors: Divide at/Semicolon
 IPCR Classifications
 IPCR Classifications: Divide at/Semicolon
 Jurisdiction
 Kind
 Lens ID
 Non-Patent Citations
 Priority Numbers
 Priority Numbers: Divide at/Semicolon
 Publication Date
 Publication Number

New Field: **Applicants: Divide at/Semicolon (Cleaned) (Cleaned)**

Confirm Changes
 Verify Matches w/ another Field

Setup Verification

Cleanup's Performance Goal
 Faster

Resolving Indeterminate Group Tags
 Mark neutral (most conservative)
 Based on record count (less conservative)
 Mark include (least conservative)

Look in: **Fuzzy**

Recent Places: Desktop, Libraries, Computer, Network

Name

- BritishAmericanSpelling.fuz
- BritishAmericanSpellingStem.fuz
- General.fuz
- Organization Names (ignore dept).fuz
- Organization Names.fuz
- Person Names.fuz
- Person Names-Aggressive.fuz
- Person Names-Very Aggressive.fuz

Choose Field

Choose the field you want to use to verify matches in List Cleanup. This means that if list cleanup decides that terms A and B match, it will confirm this by finding every record that contains term A. Then, for each term in this field in those records, it tries to see if there is a record that term B is in that also contains that term. If so it keeps the match. If not it rejects the match.

CPC Classifications
 CPC Classifications: Divide at/Semicolon
 DOI(s)
 DOI(s): Divide at/Semicolon
 Extended Family Size
 Full Text
 Inventors
Inventors: Divide at/Semicolon
 IPCR Classifications
 IPCR Classifications: Divide at/Semicolon
 Jurisdiction
 Kind
 Lens ID
 Non-Patent Citations

Use Fuzzy Match

C:\Program Files (x86)\VantagePoint\Fuzzy

OK **Cancel**

File name: **Organization Names**
 Files of type: **Fuzzy Matching Files (*.fuz)**

Organization Names (Patent Assignees, Corporate Source) – 68% match; At least 51% agreement b/w matching terms
 Uses Stemming, Ignore List: OrgIgnore.dat

Setup verification: Inventors: Divide at Semicolon. OK

Cleanup Confirm

Cleanup Sets Detail Window

Item Name	Num Records	Display
OBSHESTVO S OGRANICHENNOJ OTVETSTVENNOSTJU PARAFARM	11	<input type="radio"/> All Items
OBSHESTVO S OGRANICHENNOJ OTVETSTVENNOSTJU PARAFARM	2	<input checked="" type="radio"/> Combined Items
OBSHESTVO S OGRANICHENNOJ OTVETSTVENNOSTJU PARAFARM	9	<input type="radio"/> Custom Set of Items
BAODING LONG TENG SPORTS GOODS CO LTD	10	
BAODING LONG TENG SPORTS GOODS	3	
BAODING LONG TENG SPORTS GOODS CO LTD	7	
FEDERAL NOE G BJUDZHETNOE OBRAZOVATEL NOE UCHREZHDENIE VYSSHEGO PROFESSIONAL NOGO OBRAZOVANI...	4	
FEDERAL NOE G BJUDZHETNOE OBRAZOVATEL NOE UCHREZHDENIE VYSSHEGO PROFESSIONAL NOGO OBRAZOV...	2	
FEDERAL NOE G OBRAZOVATEL NOE UCHREZHDENIE VYSSHEGO PROFESSIONAL NOGO OBRAZOVANIJA STAVROP...	2	
BEIJING HANGTIAN KEYI TECHNOLOGY CO LTD	3	
BEIJING AEROSPACE KEYI TECHNOLOGY CO LTD	1	
BEIJING HANGTIAN KEYI TECHNOLOGY CO LTD	2	
BURDETT MICHAELL	3	
BURDETT MICHAELL	2	
BURDETT MICHAEL LEE	1	
DINGZHOU CITY LISHENG SPORTS GOODS CO LTD	3	
DINGZHOU CITY LISHENG SPORTS GOODS CO LTD	2	
DINGZHOU LISHENG SPORTING GOODS CO LTD	1	
PARROT SA	3	
PARROT SA	2	
PARROT SA	1	
HODGE MALCOLM H	2	
HODGE MALCOLM H	1	
MALCOLM H HODGE	1	
LANXI HONGXIANG BIO TECHNOLOGY CO LTD	2	
LANXI HONGXIANG BIO TECHNOLOGY CO LTD	1	
LANXI HONGXIANG BIOLG SCIENCE AND TECHNOLOGY CO LTD	1	
VĀZKUMNĀ ĀISTAV MELIORACĀ A OCHRANY PĀRDY	2	
VĀZKUMNĀ ĀISTAV MELIORACĀ A OCHRANY PĀRDY	1	
ZKUMNĀ ĀISTAV MELIORACĀ A OCHRANY PĀRDY V	1	

Save as Thesaurus

Accept

Cancel

Looks Good > Save As Thesaurus > Choose existing > Merge the
Accept

List Cleanup

~Raw Record

Applicants

Applicants: Divide at/Semicolon

Applicants: Divide at/Semicolon (Cleaned)

Applicants: Divide at/Semicolon (Cleaned) (Cleaned)

Application Date

Application Number

Cited Count

CPC Classifications

CPC Classifications: Divide at/Semicolon

DOI(s)

DOI(s): Divide at/Semicolon

Extended Family Size

Full Text

Inventors

Inventors: Divide at/Semicolon

IPC Classifications

IPC Classifications: Divide at/Semicolon

Jurisdiction

Kind

Lens ID

Non-Patent Citations

Priority Numbers

Priority Numbers: Divide at/Semicolon

Publication Date

New Field: Applicants: Divide at/Semicolon (Cleaned) (Cleaned)

Confirm Changes

Verify Matches w/ another Field

Setup Verification

Cleanup's Performance Goal

Faster More Complete

Resolving Indeterminate Group Tags

Mark neutral (most conservative)

Based on record count (less conservative)

Mark include (least conservative)

Look in: Fuzzy

Name	Date modified	Type
BritishAmericanSpelling.fuz	03/07/2009 13:21	FUZ File
BritishAmericanSpellingStem.fuz	03/07/2009 13:21	FUZ File
General.fuz	03/07/2009 13:21	FUZ File
Organization Names (ignore dept).fuz	03/07/2009 13:21	FUZ File
Organization Names.fuz	03/07/2009 13:21	FUZ File
Person Names.fuz	15/07/2013 15:17	FUZ File
Person Names-Aggressive.fuz	29/04/2013 12:22	FUZ File
Person Names-Very Aggressive.fuz	03/07/2009 13:21	FUZ File

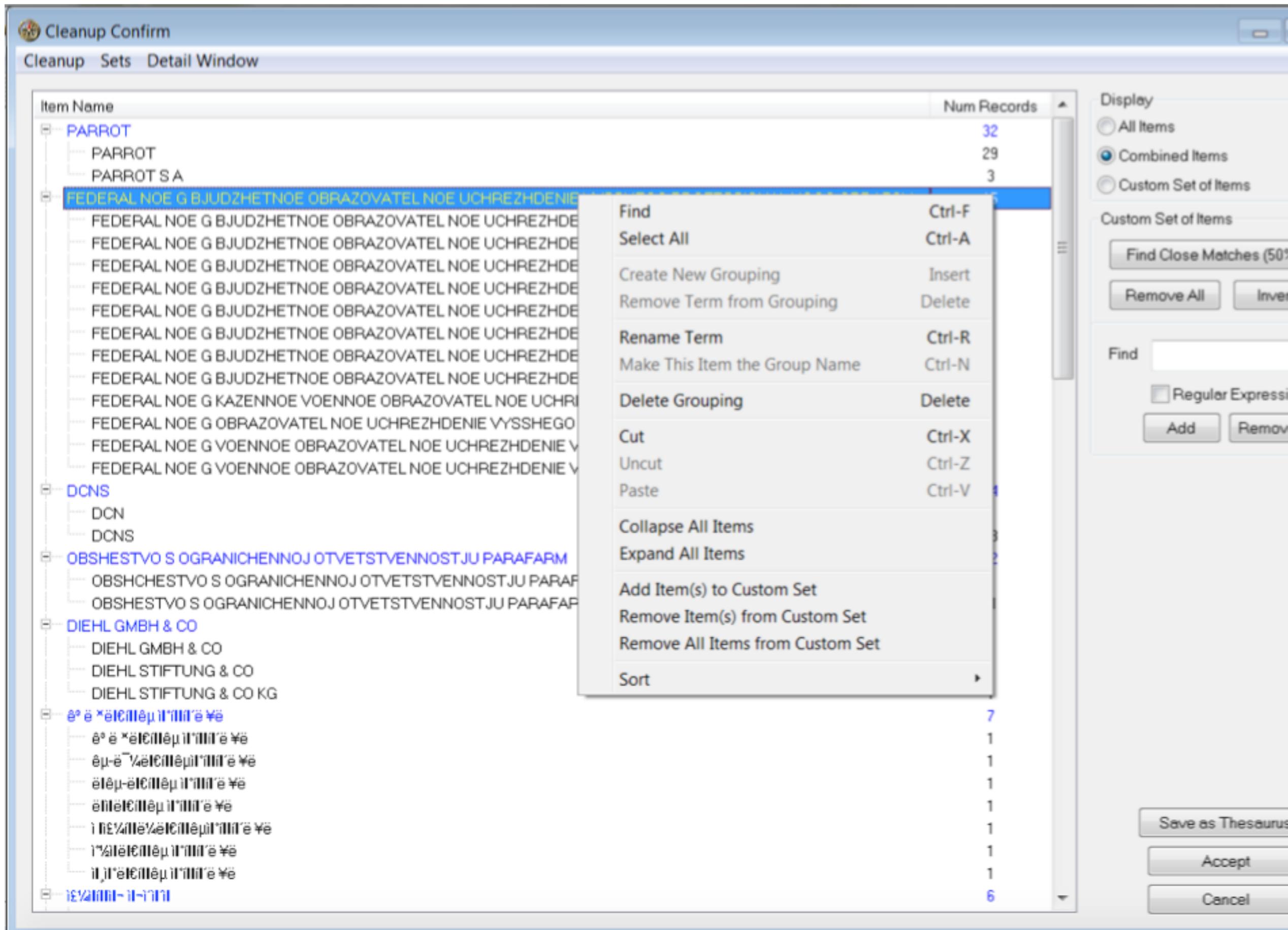
File name: Organization Names

Files of type: Fuzzy Matching Files (*.fuz)

Use Cancel

Organization Names (Patent Assignees, Corporate Source) – 68% match; At least 51% agreement b/w matching terms
Uses Stemming, Ignore List: OrgIgnore.dat

Cleanup with no verification



If uncertain - delete the grouping

Cleanup Confirm

Cleanup Sets Detail Window

Item Name	Num Records	Display
3M INNOVATIVE PROPERTIES CO	1	<input checked="" type="radio"/> All Items
AUEV ADVANCED UNMANNED ELECTRIC VEHICLE	1	<input type="radio"/> Combined Items
AAI CORP	1	<input type="radio"/> Custom Set of Items
ABDULRAB HABIB	1	
ABL IP HOLDING LLC	1	
ABRAMSON ROD	1	
ACT LIGHTING DESIGN	4	
ADACEL INC	1	
ADAMS WILLIAM J C	1	
ADCIS S R O	2	
ADIDAS AG	1	
ADVANCED LIQUID LOGIC INC	1	
ADVANCED TECH MATERIALS	1	
ADVANTEN SOC PAR ACTIONS SIMPL	1	
AEHROKON AOZT	3	
AERODRONES SARL	1	
AEROSPACE SCI & IND SHENZHEN	1	
AEROSPATIALE	4	
AEROSPATIALE MATRA	2	
AIRBUS DEFENCE & SPACE GMBH	2	
AIRBUS DS SAS	2	
AIRBUS HELICOPTERS DEUTSCHLAND GMBH	2	
AIRINOV	1	
AIRMAP INC	1	
AIROBOTICS LTD	1	
AIWA CO	1	
AIX MARSEILLE UNIVERSITE	1	
AJU CORP CO LTD	1	
AKADEMIJA	4	
AKTSIONERNOE OBSHCHESTVO KB PRIBOROSTROENIJA IM AKADEMIKA A G SHIPUNOVA	1	
AKTSIONERNOE OBSHCHESTVO KONTSERN RADIOSTROENIJA VEGA	1	
AKTSIONERNOE OBSHCHESTVO NI INZH INST AO NIII	1	
ALBINGER BERND	1	
ALCATEL LUCENT	2	
ALDERUCCI DEAN P	1	

Save as Thesaurus

Accept

Cancel

Final step - Choose All items, review, create groups, thes, accept

Facet / Filter Undo / Redo 24

Refresh Reset All Remove All

Applicants change
1312 choices Sort by: name count Cluster

NA 63
PARROT 29
THALES SA 27
OBSHESTVO S OGRANICHENNOJ OTVETSTVENNOSTJU PARAFARM 15
IDEAL TOY CORP 14
DCNS 13
BAODING LONG TENG SPORTS GOODS CO LTD 10
MATSUSHITA ELECTRIC IND CO LTD 10
US NAVY 10
BOEING CO 8
G MASHINOSTROITEL NOE KB RADUGA IM A JA BEREZNJAKA AOOT 8
HYUNDAI MOTOR CO LTD 8
ALMUHAIRBI EIDA 7
ALSHDAIFAT WASFI 7
DIEHL GMBH & CO 7
EUROCOPTER FRANCE 7

1907 rows
Show as: rows records

Applicants Inven

	# Records	# Instances	Applicants_cleaned_useme:remove
1	32	32	PARROT
2	27	27	THALES SA
3	16	16	OBSHESTVO S OGRANICHENNOJ OTVET
4	14	14	DIEHL GMBH & CO
5	14	14	IDEAL TOY CORP
6	13	13	DCNS
7	10	10	BAODING LONG TENG SPORTS GOODS C
8	10	10	EADS (various locations)
9	10	10	MATSUSHITA ELECTRIC IND CO LTD
10	10	10	US NAVY
11	8	8	BOEING CO
12	8	8	G MASHINOSTROITEL NOE KB RADUGA I
13	8	8	HYUNDAI MOTOR CO LTD
14	7	7	ALMUHAIRBI EIDA
15	7	7	ALSHDAIFAT WASFI
16	7	7	EUROCOPTER FRANCE

The result 1396 - 1287 names (109 merged).
 Open refine was good but less accurate on capture... and we could not check on other fields.

Home > 10679 (5884 families) results for ""drone" || "drones""

Results Analysis

Inventors

Highcharts.com

Jurisdictions

Highcharts.com

Feedback

Refine Search

[Clear All](#)

- Dates
- Jurisdictions
- Inventors
- Owners (US)
- Applicants
- Cited Authors
- Cited Articles (PubMed)
- Cited Articles (CR)
- Document Families
- Classifications

"drone" || "drones"

[Refine](#) [New Search](#) [⚙️](#)

+ drone_drones_ft_5884_families - 5884
- Private, Updated: Sep 21st 2016, 20:17

[View](#) [Edit](#) [Share](#) [Create Collection](#)

[File](#) [List](#) [Grid](#)

[Sort by Rank](#)

Document	Sort by Rank
<input type="checkbox"/> Bagpipe Drones Published: Oct 2, 2002 Family: 8 Cited: 0 Info: 🔗 Applicant: Hepburn Ralph Robertson	EP 0998739 B1 Doc type: Granted Patent ID: lens.org/025-809-394-828-584
<input type="checkbox"/> System And Method For Automated Traffic Management Of Intelligent Unmanned Aerial Vehicles Published: Apr 14, 2016 Family: 2 Cited: 0 Info: 🔗 Applicant: Postrel Richard	WO 2016/057098 A2 Doc type: Patent Application ID: lens.org/053-718-473-478-449
<input type="checkbox"/> Airborne Drone Formation Control System Published: May 28, 1996 Family: 1 Cited: 77 Info: 🔗 Owner: Honeywell Inc	US 5521817 A Doc type: Granted Patent ID: lens.org/020-926-965-964-522