

Universidade de São Paulo
Sistema Integrado de Bibliotecas
Departamento Técnico

**TUTORIAL PARA ANÁLISE BIBLIOMÉTRICA
DA PRODUÇÃO CIENTÍFICA DA UNIDADE
COM O USO DA WEB OF SCIENCE
E DO VANTAGEPOINT 8.0**

Eidi Raquel Franco Abdalla – FM
Girlei Aparecido de Lima - FZEA

São Paulo
2014

UNIVERSIDADE DE SÃO PAULO

Reitor: Prof. Dr. Marco Antonio Zago
Vice-Reitor: Prof. Dr. Vahan Agopyan

**SISTEMA INTEGRADO DE BIBLIOTECAS DA USP
DEPARTAMENTO TÉCNICO**

Mariza Leal de Meirelles Do Coutto

GRUPO DE ESTUDOS BIBLIOMÉTRICOS APLICADOS DO SIBiUSP

Elisabeth Adriana Dudziak – DT/SIBi - (Coordenação)
Sibele Fausto – DT/SIBi
Roseli Matsuda – DT/SIBi
Eidi Raquel Franco Abdalla – FM
Stela do Nascimento Madruga - FMVZ
Rubenildo Oliveira da Costa – FO
Girlei Aparecido de Lima - FZEA
Irene Lucinda – ICMC
Gracielle Pepe – IFSC
Daniel Jorge Caetano - DT/SIBi
Ricardo Amaral de Faria – DT/SIBi
Tarcisio Pereira – DT/SIBi

© 2014 SIBiUSP

Sistema Integrado de Bibliotecas
Universidade de São Paulo
Rua da Biblioteca, s/n
05508-050 – São Paulo – SP
<http://www.bibliotecas.usp.br>
e-mail: atendimento@sibi.usp.br

Publicação licenciada sob uma licença:
CC BY-NC-ND

INTRODUÇÃO

O presente tutorial visa fornecer orientações acerca do tratamento bibliométrico automatizado a partir de registros recuperados na base de dados Web of Science (WoS) e com o uso de software de mineração de dados tecnológicos.

A WoS é uma base de dados bibliográficos de abrangência internacional e multidisciplinar, mantida pela Thomson Reuters (antigo ISI – *Institute for Scientific Information*). Permite a recuperação de artigos e trabalhos de congressos publicados nos mais importantes periódicos internacionais, com informações sobre os trabalhos que os citaram, link para texto completo, exportação de registros para gerenciadores de referências, entre outras facilidades.

O VantagePoint é um software de mineração de dados tecnológicos que facilita as análises bibliométricas.

O objetivo deste tutorial é capacitar os bibliotecários e técnicos do Sistema Integrado de Bibliotecas (SIBiUSP) na utilização do software VantagePoint (VP), mais precisamente na coleta e tratamento dos registros bibliográficos indexados nessa base de dados, para atender à comunidade acadêmica.

O software permite identificar o perfil dos registros bibliográficos, a evolução temporal, os principais autores, áreas de pesquisa e assuntos, principais instituições e países, dentre outros indicadores de forma rápida e eficaz.

MÉTODOS

Pesquisa na base de dados

Para recuperar registros de uma determinada Unidade USP na Web of Science, inicialmente se fez necessário definir as principais formas de se escrever o nome da Unidade, levando-se em conta nome por extenso, nome abreviado, sigla da Unidade, dos departamentos e até mesmo dos laboratórios, e também o idioma em que essas palavras são escritas. Para isto, recomendamos que cada Unidade faça buscas exploratórias na WoS, no campo *Endereço*, a fim de identificar palavras e termos mais utilizados relacionados à Unidade USP.

Outra possibilidade para recuperar a Produção Científica das Unidades USP é efetuar buscas, no campo *Autor*, pelo nome científico dos pesquisadores da Unidade. Para isto, recomenda-se consultar o Currículo Lattes de todos os pesquisadores da Unidade. E neste caso, os nomes científicos seriam digitados no campo *Autor* separados entre si por OR.

Há inúmeras possibilidades de busca na WoS e, quanto mais familiarizado o profissional com os recursos existentes, mais precisa será a resposta. Pode-se usar, por exemplo, o ResearchID do pesquisador.

Para possibilitar uma maior recuperação de registros, recomenda-se usar os operadores booleanos AND, OR, NOT, SAME e NEAR, e o operador de truncamento, que na WoS é representado pelo asterisco (*).

É muito importante documentar a data e a hora da coleta de dados, pois a base é dinâmica e possui atualizações diárias.

Para dar início às buscas na WoS, acesse <http://www.sibi.usp.br/> > clique em Bases de Dados > Multidisciplinar > Web of Science, ou diretamente no seguinte link:

http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?highlighted_tab=WOS&product=WOS&last_prod=WOS&search_mode=GeneralSearch

Exemplo de expressão de busca:

WEB OF SCIENCE™

Pesquisa Principal Coleção do Web of Science™ Minhas ferramentas

Bem-vindo ao novo V

Pesquisa Básica

usp OR univ sao paulo OR sao paulo univ*

Visualizar lista de abreviações

+ Adicionar outro campo

Após efetuar a busca, os registros recuperados deverão ser exportados para posterior análise. O procedimento de envio de resultados é descrito a seguir. Como a Web of Science permite salvar no máximo 500 registros de cada vez, caso haja uma grande quantidade de registros, consulte o [Tutorial de Instalação do iMacros para o navegador Firefox: automatizando o download de registros bibliográficos da base Web of Science](#).

Envio dos resultados

Após selecionar todos os registros recuperados, clique em “Salvar em outros formatos de arquivo”. Aparecerá a seguinte tela:

Enviar para arquivo

Número de registros: Todos os registros na página
 Registros até

Gravar conteúdo: Registro completo

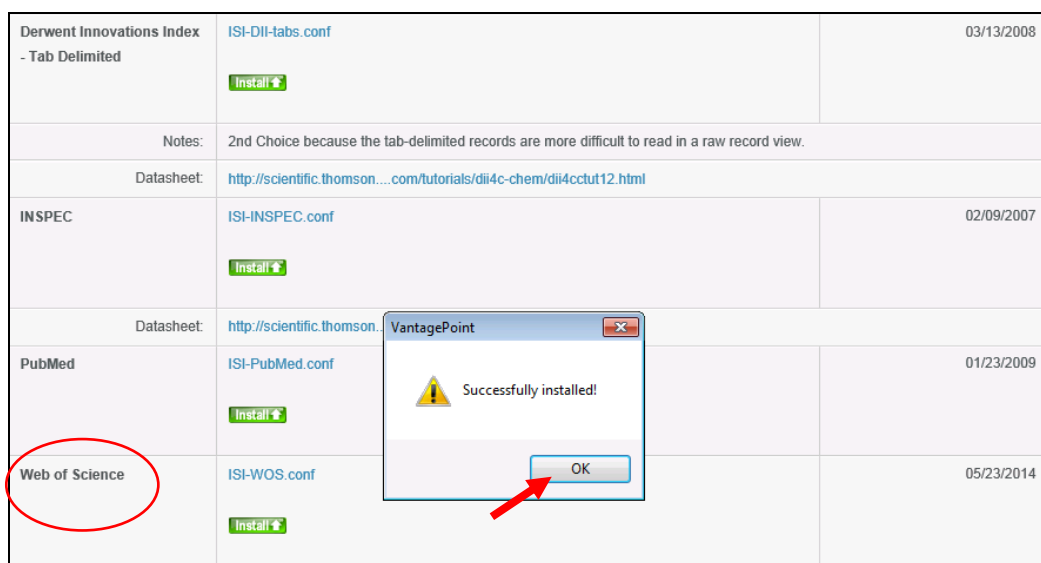
Formato de arquivo: Texto sem formatação

Enviar Cancelar

Conforme a ilustração, selecione “**Registro completo**”, e “**Texto sem formatação**” e em seguida “**Enviar**”. O arquivo *savedrecs.txt* será salvo na pasta Downloads (na raiz, em usuários). Em seguida, dá-se início ao tratamento dos mesmos no software VantagePoint.

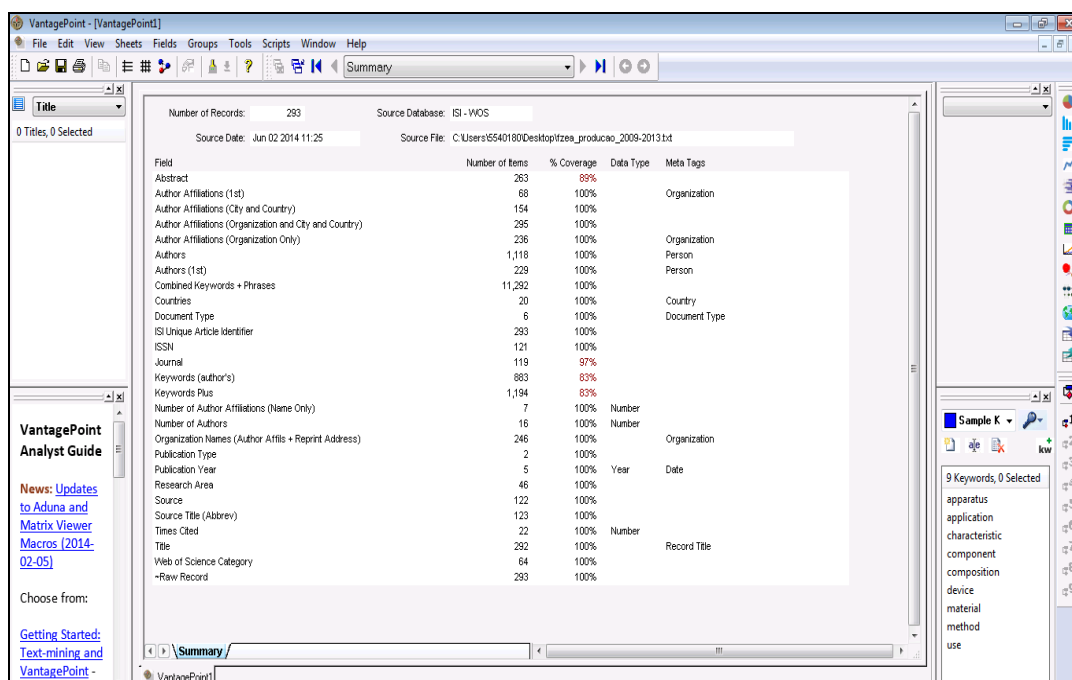
Importação no software de *tech mining*

Ao abrir o VP, clicar em *Import Raw Data File (Use Import Filter)*, clicar em *OK*. Em seguida, *Select Files*, procure o arquivo *savedrecs.txt* salvo em seu computador (caso o resultado da busca tenha sido salvo em mais de um arquivo .txt, deve-se selecionar todos os arquivos e importar tudo de uma única vez), e depois clique em *Next*. Aparecerá uma tela para escolher o filtro. Selecione ISI-WOS. Caso seja a primeira vez e não esteja disponível a opção de filtro da WoS, clique em *Select New Filter Directory*. Clique em: *Arquivos de Programas (x86) > VantagePoint > Import Filters*. Caso o filtro da WoS não esteja salvo em seu computador, clique em *Help > Visit Download Pages*. Uma vez na página de Downloads, selecione *Import Filters* e clique em *ISI-WOS.conf*, conforme ilustração abaixo:



Após a instalação do filtro da WoS, aparecerá na lista de filtros do VP, conforme descrito no início do parágrafo anterior. O uso do filtro se faz necessário porque importa automaticamente os dados de acordo com o padrão de metadados adotado pela base.

Após a importação dos registros para o VP, salve o trabalho (o usuário deve salvar periodicamente suas ações desenvolvidas no VP. Enfim, após abrir o arquivo .txt no VantagePoint e selecionar o filtro, será apresentado um Sumário, com todos os campos, número de itens, porcentagem de cobertura em cada campo, tipo de dado e meta tags, conforme a figura a seguir.



VantagePoint - [VantagePoint]

Number of Records: 293 Source Database: ISI - WOS
Source File: C:\Users\ES40180\Desktop\pizaa_producao_2009-2013.txt
Source Date: Jun 02 2014 11:25

Field	Number of Items	% Coverage	Data Type	Meta Tags
Abstract	263	89%		
Author Affiliations (1st)	68	100%		Organization
Author Affiliations (City and Country)	154	100%		
Author Affiliations (Organization and City and Country)	295	100%		
Author Affiliations (Organization Only)	236	100%		Organization
Authors	1,118	100%		Person
Authors (1st)	229	100%		Person
Combined Keywords + Phrases	11,292	100%		
Countries	20	100%		Country
Document Type	6	100%		Document Type
ISI Unique Article Identifier	293	100%		
ISSN	121	100%		
Journal	119	97%		
Keywords (author's)	883	83%		
Keywords Plus	1,194	83%		
Number of Author Affiliations (Name Only)	7	100%	Number	
Number of Authors	16	100%	Number	
Organization Names (Author Affil + Reprint Address)	246	100%		Organization
Publication Type	2	100%		
Publication Year	5	100%	Year	Date
Research Area	46	100%		
Source	122	100%		
Source Title (Abbrev)	123	100%		
Times Cited	22	100%	Number	
Title	292	100%		Record Title
Web of Science Category	64	100%		
-Raw Record	293	100%		

VantagePoint Analyst Guide
News: Updates to Aduna and Matrix Viewer
Macros [2014-02-05]
Choose from:
Getting Started: Text-mining and VantagePoint -

Sample K
9 Keywords, 0 Selected
apparatus
application
characteristic
component
composition
device
material
method
use

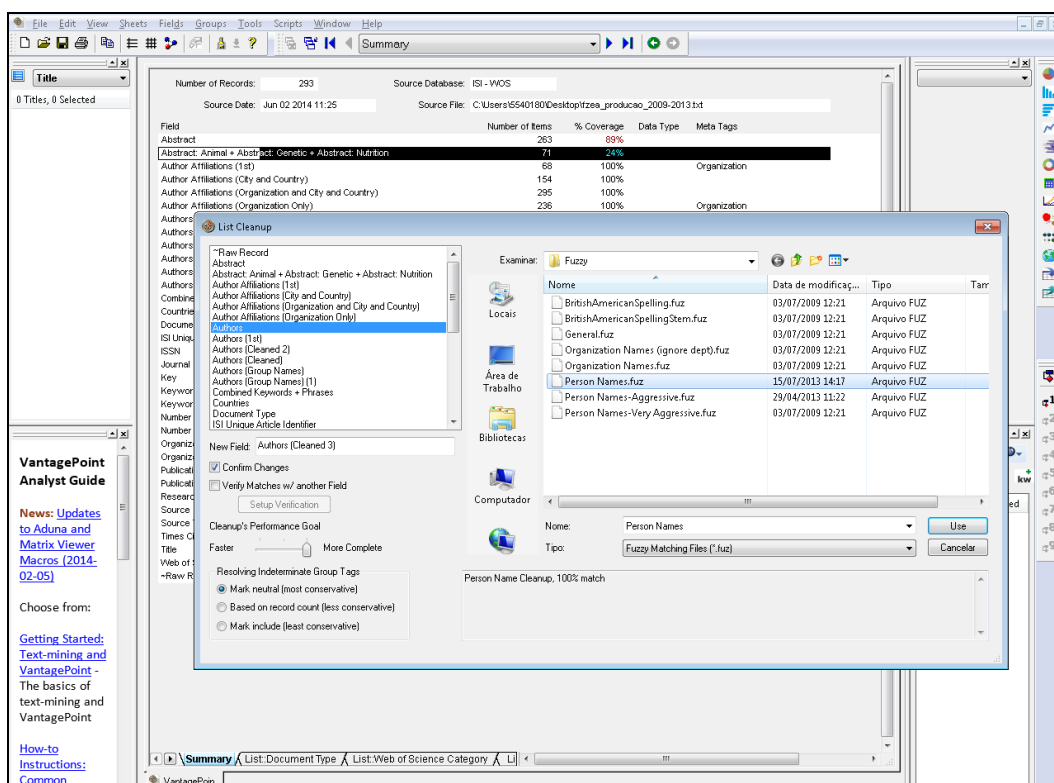
Análise de dados

A partir deste Sumário, é possível gerar inúmeros indicadores, seja por meio de listas, matrizes ou mapas.

Observem que alguns campos não possuem a cobertura 100% (coluna *Coverage*) isto é, nem todos os registros possuem o campo *Journal*, pois há outras tipologias de documentos que não contemplam este campo (*Proceedings Paper*, *Letter*, etc). Também não são todos os registros que possuem *keyword* do autor, *keyword plus*, nem *abstract*. No entanto, isso não impede a análise, e a maioria dos campos apresenta 100% de cobertura podendo ser analisada em sua totalidade.

Ao dar um duplo clique sobre o campo *Authors*, por exemplo, o VP gera uma lista de autores. Esta lista pode ser ordenada alfabeticamente ou pela coluna *Records*, que ordena pelo número de registros atrelados ao autor. Para ordenar, basta colocar o cursor no início da coluna e dar um duplo clique.

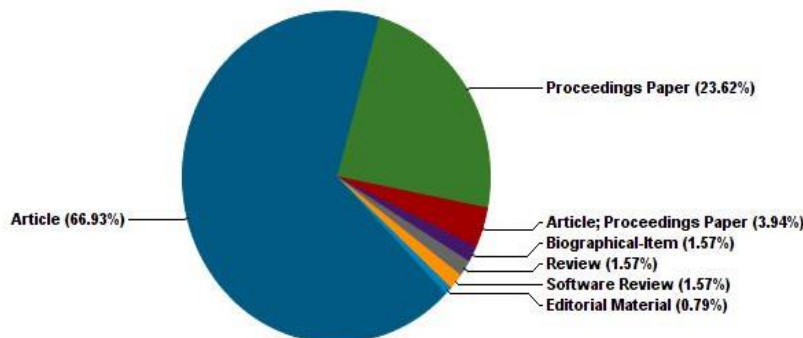
No entanto, alguns campos precisam ser mais bem estruturados e para isto há algumas ferramentas que devem ser utilizadas. O campo *Authors*, por exemplo, traz algumas variações na grafia do nome dos autores. Para normalizar estas variações, há necessidade de se efetuar uma limpeza nos dados. Clicar em *Fields > List cleanup > selecione o campo Authors > Dentro da pasta Fuzzy, selecione Person Name.fuz* (conforme a ilustração abaixo).



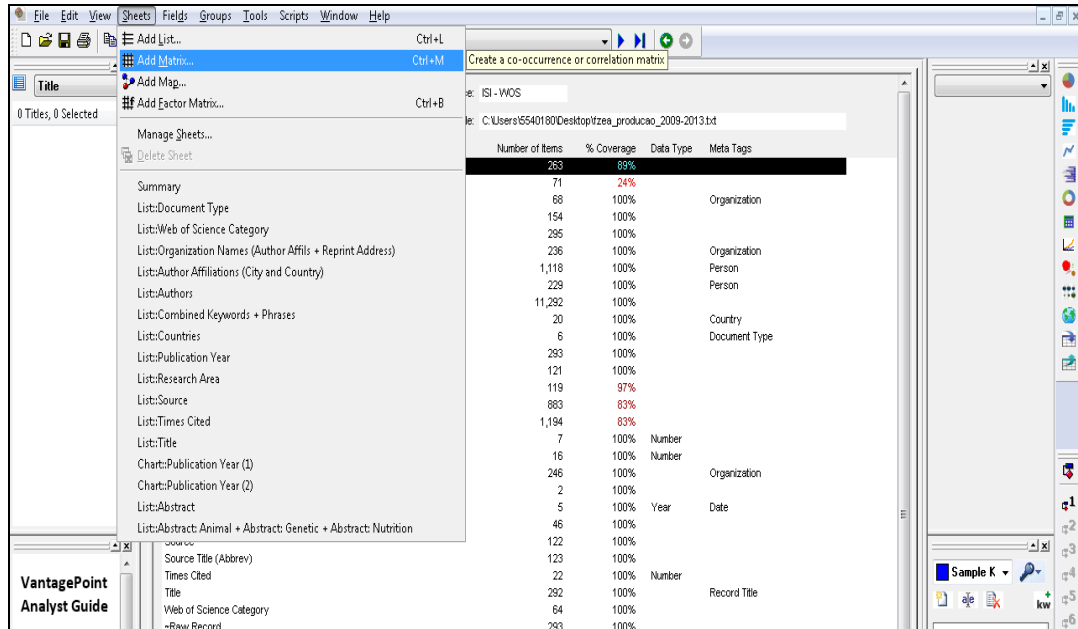
Ao clicar em *Use*, será aberta uma nova janela chamada *Cleanup Confirm*. Clicar em *All Items* e arrastar as variações de nomes para o nome normalizado. Após finalizar o agrupamento, clique em *Accept* e em seguida *Save as Thesaurus* – funcionalidade que é muito útil para análises futuras (ver sobre Thesaurus no [Manual do VantagePoint](#)). Com este procedimento é gerada uma nova lista de autores, limpa e normalizada. Salve a pesquisa com a extensão **.vpt**.

Este mesmo método pode ser aplicado para o campo *Organization Name*, para agrupar as variações dos nomes das instituições.

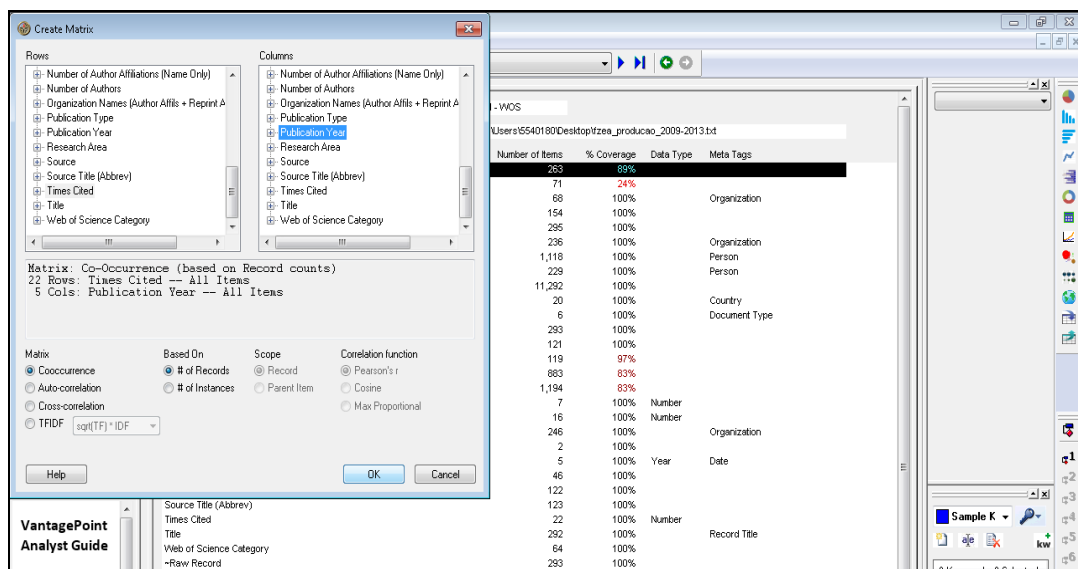
Dentre muitos indicadores possíveis de se extrair da WoS, um dos mais utilizados é a tipologia de documentos. Após gerar a lista de *Document Type*, selecionar as linhas e colunas e clicar na barra lateral direita em *Pie Chart* (primeiro ícone, em formato de pizza).



O VP possibilita a criação de inúmeras matrizes. Clicar em *Sheets > Add Matrix*. É possível criar muitos relacionamentos entre as listas, a fim de gerar importantes e variados indicadores.



No exemplo abaixo temos uma matriz de “Citações recebidas” com “Ano de publicação”.

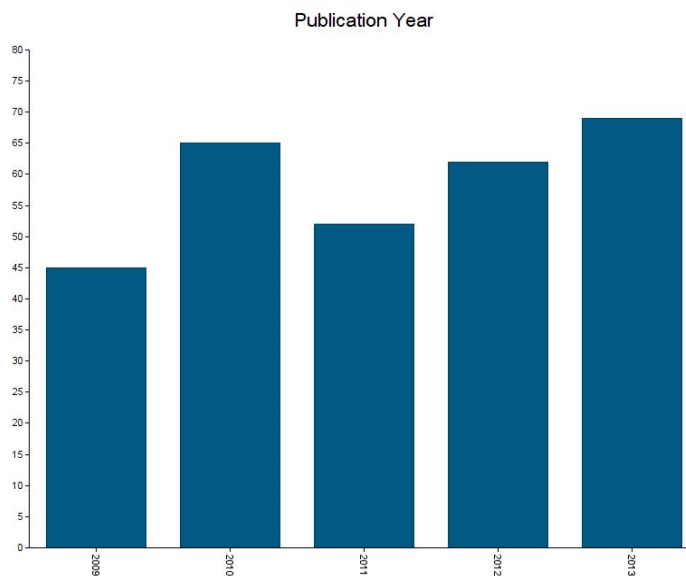


Neste momento, é preciso definir se a matriz será de coocorrência, autocorrelação ou crosscorrelação. São três tipos diferentes de matrizes geradas pelo VP, e cada uma oferece análises distintas (para mais detalhes sobre cada tipo de matriz, ver o Manual do VantagePoint). Neste exemplo, foi selecionada a opção *Cooccurrence*, e em seguida clicado OK.

Publication Year	# Records	1	2	3	4	5
1	132	0	12	39	16	23
2	57	1	8	7	13	21
3	8	5	4	2	2	0
4	28	2	3	8	4	8
5	5	9	3	2	0	0
6	10	4	2	6	1	1
7	2	21	2	0	0	0
8	10	7	1	5	3	1
9	3	6	1	2	0	0
10	6	10	1	3	1	1
11	2	14	1	1	0	0
12	4	11	1	1	0	2
13	3	12	1	2	0	0
14	2	13	1	1	0	0
15	1	19	1	0	0	0
16	1	25	1	0	0	0
17	1	28	1	0	0	0
18	1	47	1	0	0	0
19	11	3	2	3	5	1
20	4	8	1	1	2	0
21	1	27	0	0	1	0
22	1	17	1	0	0	0

É possível fazer muitas combinações de dados por meio das matrizes. Combinar Ano x assunto, para saber a evolução das pesquisas no decorrer dos anos acerca de um assunto. Ou combinar áreas de pesquisa com país, para conhecer os países que mais desenvolvem pesquisas numa determinada área, entre muitas outras possibilidades.

Outro importante indicador é a evolução temporal, que permite conhecer os dados quantitativos no decorrer dos anos. Ao gerar a lista *Publication Year*, clique nas opções de gráfico, no lado direito da tela e gere uma ilustração de forma rápida.



Quanto às citações recebidas, basta clicar em *Times Cited*, para gerar lista de quantas vezes cada registro bibliográfico foi citado. O exemplo abaixo mostra que 132 artigos não receberam nenhuma citação. Cinquenta e sete artigos receberam 1 citação, e assim sucessivamente. Interessante notar que teve um artigo recebeu 47 citações.

	# Records	# Instances	Times Cited
1	132	132	0
2	57	57	1
3	28	28	2
4	11	11	3
5	10	10	7
6	10	10	4
7	8	8	5
8	6	6	10
9	5	5	9
10	4	4	11
11	4	4	8
12	3	3	12
13	3	3	6
14	2	2	21
15	2	2	14
16	2	2	13
17	1	1	47
18	1	1	29
19	1	1	27
20	1	1	25
21	1	1	19
22	1	1	17

Para saber quais são os periódicos onde as pesquisas estão sendo publicadas, basta dar um duplo clique em *Source* e será gerada uma lista com os títulos dos periódicos. Muitas vezes se faz necessário normalizar esta lista. Há possibilidade de ranqueamento dos títulos, definindo os *Top 10*, *Top 20* ou mais. Para criar grupo com os 10 principais títulos, por exemplo, selecione as dez primeiras linhas e clique com o botão direito. Em seguida, clique em *Add Selection to Group*, dê um nome para este grupo e em seguida OK, conforme a ilustração abaixo.

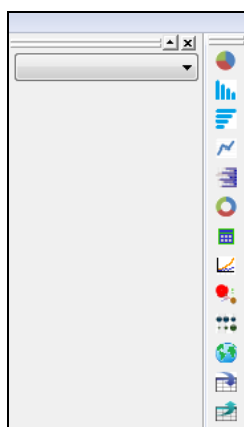
The screenshot shows a software window titled "List:Source" with a menu bar (Fields, Groups, Tools, Scripts, Window, Help) and a toolbar. The main area contains a table with the following columns: #, # Records, # Instances, and Source. The table lists 47 records, with the first 10 highlighted in black. A dialog box titled "Add items" is open on the right, with the text "Add selected items to group:" and a large empty text area. Below the text area, there is a "New group:" label and a text input field containing "Top 10". A red arrow points to the input field. There are "OK" and "Cancel" buttons at the bottom of the dialog box.

#	# Records	# Instances	Source
1	26	26	REVISTA BRASILEIRA DE ZOOTECNIA-BRAZILIAN JOURNAL OF ANIMAL SCIENCE
2	17	17	PESQUISA VETERINARIA BRASILEIRA
3	15	15	PLACENTA
4	11	11	CENCIA RURAL
5	10	10	SCIENTIA AGRICOLA
6	9	9	JOURNAL OF DAIRY SCIENCE
7	9	9	REPRODUCTION IN DOMESTIC ANIMALS
8	8	8	JOURNAL OF FOOD ENGINEERING
9	7	7	ARQUIVO BRASILEIRO DE MEDICINA VETERINARIA E ZOOTECNIA
10	7	7	REPRODUCTION FERTILITY AND DEVELOPMENT
11	6	6	ITALIAN JOURNAL OF FOOD SCIENCE
12	5	5	REVISTA BRASILEIRA DE CIENCIA DO SOLO
13	5	5	THERIOGENOLOGY
14	4	4	ANIMAL REPRODUCTION SCIENCE
15	4	4	INTERNATIONAL JOURNAL OF FOOD SCIENCE AND TECHNOLOGY
16	4	4	JOURNAL OF CHEMICAL AND ENGINEERING DATA
17	4	4	MICROSCOPY RESEARCH AND TECHNIQUE
18	4	4	REPRODUCTIVE BIOLOGY AND ENDOCRINOLOGY
19	3	3	BUFFALO BULLETIN
20	3	3	CELLULAR REPROGRAMMING
21	3	3	FOOD RESEARCH INTERNATIONAL
22	3	3	INDUSTRIAL CROPS AND PRODUCTS
23	3	3	LWT-FOOD SCIENCE AND TECHNOLOGY
24	3	3	POWDER TECHNOLOGY
25	2	2	ACTA SCIENTIAE VETERINARIAE
26	2	2	ADVANCED MATERIALS FORUM V, PT 1 AND 2
27	2	2	BIOLOGY OF REPRODUCTION
28	2	2	CARBOHYDRATE POLYMERS
29	2	2	CEMENT & CONCRETE COMPOSITES
30	2	2	CIENCIA E TECNOLOGIA DE ALIMENTOS
31	2	2	CLONING AND STEM CELLS
32	2	2	FLUID PHASE EQUILIBRIA
33	2	2	FOOD CHEMISTRY
34	2	2	ICHEAP-10: 10TH INTERNATIONAL CONFERENCE ON CHEMICAL AND PROCESS ENGIN
35	2	2	INTERNATIONAL JOURNAL OF BIOMETEOROLOGY
36	2	2	JOURNAL OF AQUATIC FOOD PRODUCT TECHNOLOGY
37	2	2	JOURNAL OF CHEMICAL THERMODYNAMICS
38	2	2	JOURNAL OF POLYMER ENGINEERING
39	2	2	LIVESTOCK SCIENCE
40	2	2	NOVEL AND NON-CONVENTIONAL MATERIALS AND TECHNOLOGIES FOR SUSTAINA
41	2	2	PLOS ONE
42	2	2	RESEARCH IN VETERINARY SCIENCE
43	2	2	REVISTA BRASILEIRA DE ENGENHARIA AGRICOLA E AMBIENTAL
44	2	2	REVISTA BRASILEIRA DE FRUTICULTURA
45	2	2	ROMANIAN BIOTECHNOLOGICAL LETTERS
46	2	2	STARCH-STARKE
47	1	1	ACTA CIRURGICA BRASILEIRA

Após essa ação, o VP gera automaticamente uma nova lista mostrando o grupo dos títulos de revista selecionados como "Top 10", conforme mostrado abaixo.

	# Records	# Instances	Source	Top 10
1	26	26	REVISTA BRASILEIRA DE ZOOTECNIA-BRAZILIAN JOURNAL OF ANIMAL SCIENCE	▼
2	17	17	PESQUISA VETERINARIA BRASILEIRA	▼
3	15	15	PLACENTA	▼
4	11	11	CIENCIA RURAL	▼
5	10	10	SCIENTIA AGRICOLA	▼
6	9	9	JOURNAL OF DAIRY SCIENCE	▼
7	9	9	REPRODUCTION IN DOMESTIC ANIMALS	▼
8	8	8	JOURNAL OF FOOD ENGINEERING	▼
9	7	7	ARQUIVO BRASILEIRO DE MEDICINA VETERINARIA E ZOOTECNIA	▼
10	7	7	REPRODUCTION FERTILITY AND DEVELOPMENT	▼
11	6	6	ITALIAN JOURNAL OF FOOD SCIENCE	▼
12	5	5	REVISTA BRASILEIRA DE CIENCIA DO SOLO	▼
13	5	5	THERIOGENOLOGY	▼
14	4	4	ANIMAL REPRODUCTION SCIENCE	▼
15	4	4	INTERNATIONAL JOURNAL OF FOOD SCIENCE AND TECHNOLOGY	▼
16	4	4	JOURNAL OF CHEMICAL AND ENGINEERING DATA	▼
17	4	4	MICROSCOPY RESEARCH AND TECHNIQUE	▼
18	4	4	REPRODUCTIVE BIOLOGY AND ENDOCRINOLOGY	▼
19	3	3	BUFFALO BULLETIN	▼
20	3	3	CELLULAR REPROGRAMMING	▼
21	3	3	FOOD RESEARCH INTERNATIONAL	▼
22	3	3	INDUSTRIAL CROPS AND PRODUCTS	▼
23	3	3	LWT-FOOD SCIENCE AND TECHNOLOGY	▼
24	3	3	POWDER TECHNOLOGY	▼
25	2	2	ACTA SCIENTIAE VETERINARIAE	▼
26	2	2	ADVANCED MATERIALS FORUM V, PT 1 AND 2	▼
27	2	2	BIOLOGY OF REPRODUCTION	▼
28	2	2	CARBOHYDRATE POLYMERS	▼
29	2	2	CEMENT & CONCRETE COMPOSITES	▼
30	2	2	CIENCIA E TECNOLOGIA DE ALIMENTOS	▼
31	2	2	CLONING AND STEM CELLS	▼
32	2	2	FLUID PHASE EQUILIBRIA	▼
33	2	2	FOOD CHEMISTRY	▼
34	2	2	ICHEAP-10: 10TH INTERNATIONAL CONFERENCE ON CHEMICAL AND PROCESS ENGIN	▼
35	2	2	INTERNATIONAL JOURNAL OF BIOMETEOROLOGY	▼
36	2	2	JOURNAL OF AQUATIC FOOD PRODUCT TECHNOLOGY	▼
37	2	2	JOURNAL OF CHEMICAL THERMODYNAMICS	▼
38	2	2	JOURNAL OF POLYMER ENGINEERING	▼
39	2	2	LIVESTOCK SCIENCE	▼
40	2	2	NOVEL AND NON-CONVENTIONAL MATERIALS AND TECHNOLOGIES FOR SUSTAINA	▼

Ao criar grupos, seja com os principais títulos de periódicos, ou principais autores, ou principais assuntos/áreas de pesquisa, é possível gerar importantes gráficos, matrizes e mapas, que sintetizam e ilustram as análises bibliométricas. Ao clicar em *Scripts*, na barra de ferramentas, há inúmeras possibilidades (*Aduna Cluster Map*, *Bubble Chart*, etc.) que podem ser utilizadas. Uma forma mais rápida e simples de gerar figuras e gráficos no VP é utilizar os ícones que há no canto direito da tela do VantagePoint (ver figura abaixo).



É importante mencionar que seria impossível, na maioria dos casos, gerar ilustrações sem antes criar grupos, tendo em vista a grande quantidade de dados em cada lista. Outra limitação do VP é que não há muitas possibilidades de edição nos gráficos plotados pelo aplicativo, sendo necessário exportar as listas e matrizes geradas pelo VP para o Excel, para a edição de gráficos mais elaborados ou customizados nesse software.

CONCLUSÃO

O VantagePoint é uma ferramenta extremamente poderosa para análise bibliométrica e deve ser explorado em todos os seus recursos, fornecendo indicadores de produção, citação e ligação da Unidade, Departamento ou Grupo de Pesquisa.

Procuramos com esse tutorial apresentar algumas das possibilidades oferecidas, mas há, sem dúvida, inúmeras outras, como por exemplo, pode ser gerado um sub-dataset de um departamento ou grupo de pesquisa específico da unidade, para uma análise num nível mais aprofundado. Para finalizar, enfatizamos que nada dispensa a leitura dos guias e ajudas existentes no próprio software.