

EL MANEJO DE DATOS

Aproximación desde los estudios
de la información

Georgina Araceli Torres Vargas



Z666.73
L56M3

El manejo de datos. Aproximación desde los estudios de la información / Coordinadora Georgina Araceli Torres Vargas. - México: UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2019.

viii, 116 pp. - Colección: TECNOLOGÍAS DE LA INFORMACIÓN.

ISBN: 978-607-30-2690-1

1. Datos vinculados. 2. Minería de datos. 3. Investigación bibliotecológica.

I. Torres Vargas, Georgina Araceli, coordinadora. II. Ser.

Diseño de portada: Natalia Cristel Gómez Cabral

Primera edición, 2020

D.R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, Ciudad de México

Impreso y hecho en México

ISBN: 978-607-30-2690-1

Publicación dictaminada

2020

Contenido

Presentación.....	7
GEORGINA ARACELI TORRES VARGAS	

MINERÍA DE TEXTO Y MINERÍA DE DATOS

Identificación de los temas de investigación en los documentos científicos del Colegio de Postgraduados.	11
ÁNGEL BRAVO VINAJA	
SANTIAGO DE JESÚS MÉNDEZ GALLEGOS	
JORGE PALACIO NUÑEZ	

Minería de texto aplicada a un diagnóstico de usuarios en Ciencia y Tecnología: aprendizajes para fortalecer la investigación bibliotecológica.	31
LOURDES FERIA BASURTO	

Minería de Datos, el caso de estudio de la Biblioteca Dr. Valentín Gómez Farías de la Facultad de Medicina de la UNAM.	43
DAVID FLORES MACÍAS	
GUADALUPE VANESA CAROLINA GUTIÉRREZ HERNÁNDEZ	

SISTEMATIZACIÓN DE DATOS Y SERVICIOS DE INFORMACIÓN

Research Data Management and Libraries: Opportunities and Challenges.....	59
KRYSZYNA K. MATUSIAK	

Integración de los principios de <i>linked data</i> en el registro bibliográfico.....	75
---	----

EDER ÁVILA BARRIENTOS

Plan para el Desarrollo de la Ciencia de Datos y Big Data (PDCDBD) en la UNAM para fines académicos y administrativos.....	95
--	----

JAVIER SALAZAR ARGONZA

Identificación de los temas de investigación en los documentos científicos del Colegio de Postgraduados

ÁNGEL BRAVO VINAJA
SANTIAGO DE JESÚS MÉNDEZ GALLEGOS
JORGE PALACIO NUÑEZ
*Colegio de Postgraduados
Campus San Luis Potosí*

INTRODUCCIÓN

El Colegio de Postgraduados (ColPos) es un Centro Público de Investigación dependiente de la Secretaría de Agricultura y Recursos Hidráulicos Pesca y Alimentación (SAGARPA), creado por Decreto Presidencial en 1959 (Colegio de Postgraduados 2014). En el 2001 el ColPos se constituyó en un Centro Público de Investigación, lo que le permitió autonomía y mayor independencia presupuestal, que cuando estaba bajo control de la SAGARPA y la Secretaría de Hacienda y Crédito Público (Colegio de Postgraduados 2016). Fue concebido como una institución pública estratégica para el desarrollo social del sector agropecuario y forestal de México, a través de la formación de recursos humanos de alto nivel, para generar información científica que contribuya al desarrollo y fortalecimiento de instituciones del sector (González

Cossío 2010). Su misión es “generar, difundir y aplicar conocimiento para el manejo sustentable de los recursos naturales, la producción de alimentos nutritivos e inocuos, y el mejoramiento de la calidad de vida de la sociedad” (Colegio de Postgraduados 2016). Esta institución imparte dieciséis programas de maestría y doctorado en ciencias en sus siete Campus, ubicados en los estados de: México, Puebla, San Luis Potosí, Tabasco, Veracruz (dos campus) y Campeche, los cuales son reconocidos por el Programa Nacional de Posgrados de Calidad (PNPC) del Consejo Nacional de Ciencia y Tecnología (Conacyt) (Colegio de Postgraduados 2017). En 2017 contaba con 444 profesores (de 616 plazas académicas) con grado de doctor en ciencias, formados en universidades de todo el mundo, de los cuales 56% pertenecían en ese año al Sistema Nacional de Investigadores (Colegio de Postgraduados 2016).

Varios de sus investigadores han sido reconocidos con el otorgamiento de premios internacionales, nacionales y estatales de ciencias y artes, en las áreas de tecnología y diseño y en ciencias naturales y exactas; premios de ciencia y tecnología de los alimentos y premios Banamex, entre otros. Además, es la institución de ciencias agrícolas mexicana que cuenta con el mayor número de investigadores nacionales Nivel III en el Sistema Nacional de Investigadores (SNI) (Larqué-Saavedra 2014). Pero la influencia del ColPos no se restringe a eso, ya que es pionero en la generación de conceptos y escuelas del pensamiento en las ciencias y tecnologías agrícolas, y por haber realizado aportaciones importantes para el desarrollo agrícola entre las que destacan: resaltar la importancia de la biodiversidad en México; establecer bancos de germoplasma, e implementar estudios fundamentales de los sistemas agrícolas y de los tipos de vegetación de México. Adicionalmente los investigadores han resaltado la importancia que representan los campesinos en la domesticación y conservación de las especies, la elaboración de mapas de suelos agrícolas y su conservación, así como el establecimiento de biofábricas de hongos comestibles y agentes de control biológico (Larqué-Saavedra 2014).

Actualmente, el ColPos cuenta con 49 líneas de investigación llamadas “Líneas de Generación y/o Aplicación del Conocimiento

(LGAC-CP)", donde confluyen la especialización de las investigaciones de los profesores-investigadores que conforman el Núcleo Académico Básico (NAB) de cada programa de postgrado, que son quienes definen la naturaleza de los programas de postgrado, además, de fundamentar los proyectos de investigación de los estudiantes y facilitar de esta forma la operación de la investigación (Colegio de Postgraduados 2018).

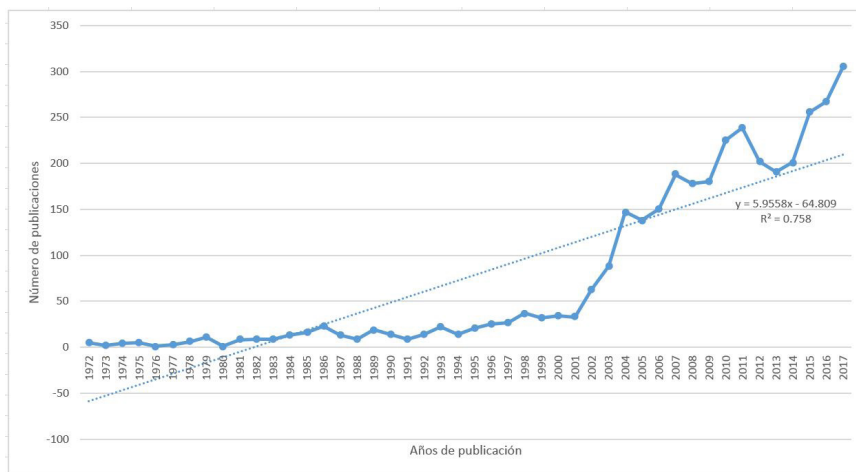
Los resultados de la actividad científica del ColPos, en sus primeros años de vida, no fue publicada en revistas internacionales indizadas en bases de datos analizadoras de la producción científica tales como el SCIE, el SSCI, contenidas en el Web de la Ciencia (WOS) y Scopus. En las bases de datos SCIE y SSCI, la primera contribución apareció hasta 1972; a partir de este momento, la publicación de contribuciones científicas hasta 1989 fue de 157 (4.54% del total publicado hasta 2017); es decir, 8.55 documentos por año. De 1990 a 2004 se publicaron 580 documentos (16.78% de lo publicado hasta 2017), esto es 38.66 por año. La mayoría de los documentos derivados de la investigación realizada en el ColPos se publicaron en revistas mexicanas, algunas de las cuales ahora aparecen listadas en el "Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología", en publicaciones seriadas del propio ColPos como "Comunicaciones en Estadística y Cómputo", "Cuadernos de Desarrollo Rural", "Comunicaciones en Socioeconomía, Estadística e Informática" y los primeros años de la revista "Agrociencia". Es en los últimos trece años que la actividad científica del ColPos se ve reflejada en las revistas de corriente principal, llamadas así por Salager-Meyer (2015) y a las revistas indizadas en las bases de datos SCIE y SSCI, ya que de 2005 a 2017 se indizaron allí 2 720 documentos del ColPos, que corresponden a 209.23 documentos por año.

En los últimos años los artículos indizados en las bases de datos SCIE y SSCI corresponden aproximadamente al 50% de la producción anual del ColPos, como lo muestran los datos estadísticos al respecto. En el año 2016 se indizaron 267 (48.72%) documentos en las bases de datos SCIE y SSCI, de 548 publicados en revistas con comité editorial reportados en el Sistema Integral de Informa-

ción Académica (SIIA) de esta institución. En 2017, la proporción subió a 51.26% (305 de 595); por lo tanto, se puede afirmar que las temáticas de investigación de la producción científica del ColPos de los últimos años que se analizan mediante minería de textos, corresponden en esta investigación al 50% de la producción total de la institución.

La producción científica del ColPos indizada en las bases de datos SCIE y SSCI, tuvo un crecimiento bajo de 1972 a 2000, pero a partir de 2001 comenzaron a indizarse un mayor número de publicaciones en las bases de datos mencionadas, presentando una tendencia creciente cada año, exceptuando un periodo entre 2011 a 2013, pero a partir de 2014 la indización de documentos volvió a crecer *Figura 1*.

Figura 1. Crecimiento de la producción científica del Colegio de Postgraduados en revistas de corriente principal de 1972 a 2017.

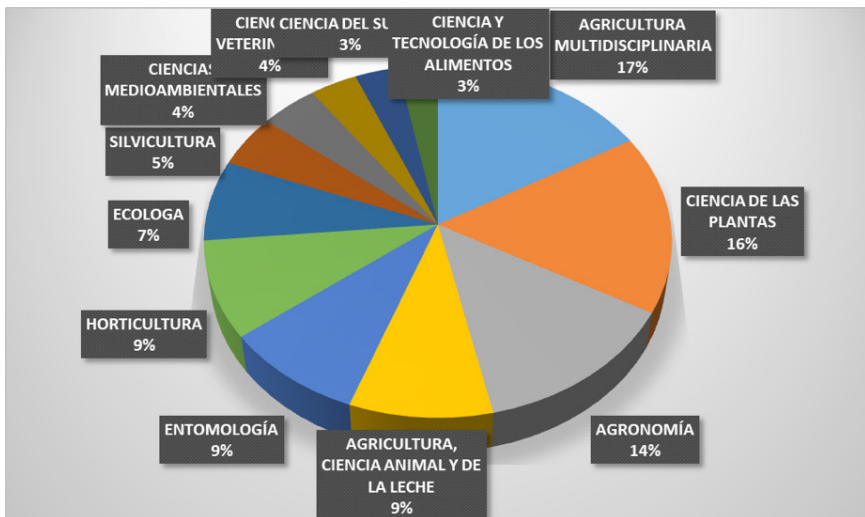


Los tipos de documentos indizados en las bases de datos SCIE y SSCI, en su mayoría, son artículos científicos (3 115, 90.17%), 199 resúmenes de congresos (5.76%), 42 editoriales (1.2%), 39 artículos *in extenso* (1.13%), 38 reseñas (1%), y el resto fueron noticias (32),

notas científicas (17), correcciones (5), biografías (3), reseñas de libros (3), cartas al editor (3), y un capítulo de libro. Vale la pena destacar la baja aportación del cuerpo académico del ColPos en la participación de libros indizados en estas bases de datos, ya que indizan principalmente revistas científicas.

En cuanto a las categorías temáticas del Web de la Ciencia, de las revistas en las que se indizaron las publicaciones del ColPos, 74% de los documentos corresponden a seis grandes temas de agricultura multidisciplinaria (17%), ciencias de las plantas (16%), agronomía (14%), ciencia animal y de la leche (9%), entomología (9%) y horticultura (9%), el resto (26%), estuvo distribuido en otras categorías tales como: ecología, silvicultura, ciencias medioambientales, ciencias veterinarias, ciencia del suelo, y ciencia y tecnología de los alimentos *Figura 2*.

Figura 2. Categorías temáticas de las revistas indizadas en el Web de la Ciencia de los documentos publicados por el Colegio de Postgraduados.



La minería de textos es el proceso de extracción de patrones o información interesante a partir de documentos de texto no estructurados (Tan 1999). En tanto que Feldman y Sanger (Feldman y Sanger 2006) lo definen como un intensivo proceso de conocimiento en el que un usuario interactúa con una colección de documentos mediante el uso de un conjunto de herramientas de análisis; mencionan además, que al igual que la minería de datos, la minería de textos busca extraer información útil de las fuentes de datos, sin embargo, en el caso de la minería de textos, las fuentes de datos son colecciones de documentos donde es posible encontrar patrones interesantes en los datos textuales no estructurados. Las aplicaciones de la minería de textos para encontrar patrones interesantes se dan principalmente, según Feldman y Sanger (2006) en áreas como la inteligencia de negocios o empresarial, el análisis de patentes, y la investigación en ciencias de la vida.

VOSviewer es un programa informático para construir y visualizar redes bibliométricas (CSTS 2018). Entre las múltiples tareas que éste puede realizar, se encuentra la minería de textos, la cual se puede realizar usando los títulos y resúmenes de los documentos. También ha sido utilizado como herramienta bibliométrica en diferentes documentos técnicos y de aplicación. Entre los documentos técnicos destacan trabajos de los creadores del programa de cómputo Ness Jan Van Eck y Ludo Waltman: “Text mining and visualization using VOSviewer” (Eck y Waltman 2007), y “VOS: A New Method for Visualizing Similarities Between Objects” (Eck y Waltman 2011). Respecto a documentos donde se aplica el análisis de textos usando Vosviewer, destaca el trabajo de Gobster (Gobster 2014) “(Text) Mining the LANDscape: Themes and trends over 40 years of Landscape and Urban Planning”.

A partir de las facilidades que proporciona VOSviewer para realizar trabajos de minería de textos usando los registros bibliográficos de diferentes bases de datos como el Web de la Ciencia y Scopus, se están realizando trabajos usando la aplicación para identificar las temáticas y tendencias de investigación, como es el caso de este trabajo, que tiene como objetivo identificar y describir

las temáticas de investigación en los documentos publicados por el personal académico del ColPos en revistas de corriente principal, lo que servirá a los tomadores de decisiones del ColPos para afianzar o reorientar la investigación científica en la institución.

METODOLOGÍA

La búsqueda de la producción científica del Colegio de Postgraduados en revistas de corriente principal se efectuó en las bases de datos SCIE y SSCI del Web de la Ciencia de la empresa Clarivate Analytics mediante la ecuación de búsqueda mostrada en la *Figura 3*, limitando los resultados desde la publicación del primer documento en 1972 hasta el año 2017.

Figura 3. Ecuación de búsqueda de la producción científica del Colegio de Postgraduados en el Science Citation index Expanded y el Social Sciences Citation index.

```
((AD=(COLEGIO POSTGRAD OR COLEGIO POSGRAD OR COLEGIO POSTGRADUADOS OR IREGEP OR
COLEGIO POSTGRAD CIENCIAS AGR OR IRENAT OR COLEGIO POSTGRAD MONTECILLO OR COLEGIO
POSGRADUADOS OR INST FITOSANIDAD OR INST RECURSOS GENET & PROD OR COLEGIO POSTGRAD
CIENCIAS AGRICOLAS OR COL POSTGRAD OR COLEGIO POSTGRAD CHAPINGO OR COLEGIO
POSTGRAD MICROBIOL EDAFOL IRENAT OR IFIT OR ISEI OR COLEGIO POSGRADUADOS
MONTECILLOS OR COLEGIO POST GRAD OR COLEGIO POSTGRAD CARRETERA MEXICO TEXCOCO OR
COLEGIO POSTGRAD CIENCIAS AGR MONTECILLO OR COLEGIO POSTGRAD CIENCIAS AGROCOLAS OR
COLEGIO POSTGRAD CIENCIAS CIENCIAS AGR OR COLEGIO POSTGRAD CONACYT OR COLEGIO
POSTGRAD EDO MEXICO OR COLEGIO POSTGRAD H CARDENAS OR COLEGIO POSTGRADOS OR
COLEGIO POSTGRADUADOS CARRETERA MEXICO TEXCOCO OR COLEGIO POSTGRADUATOS OR
COLEGIO POSTRAD OR CTR GANADERIA COLEGIO POSTGRAD OR "COLEGIO POSTGRAD, PROGRAMA
GANADERIA" OR "PROGRAMA GANADERIA, KM 36.5" OR INST RECURSOS GENET PRODIOL OR INST
RECURSOS NAT COLEGIO POSTGRAD OR INST SOCIOECON ESTADIST & INFORMAT OR IRENAT
COLEGIO POSTGRAD OR IRENAT OR IRGP OR CTR DOCUMENTAC & BIBLIOTECA, MONTECILLO OR
"CAMPUS SAN LUIS POTOSI COLEGIO POSTGRAD" OR "CAMPUS SAN LUIS POTOSI,ITURBBIDE") AND
CU=MEXICO) OR (OG=(Colegio de Postgraduados - Mexico))) NOT (AD=("UDG, COLEGIO POSTGRAD"
OR "CELULOSA & PAPEL, COLEGIO POSTGRAD"))
```

Los registros bibliográficos de los documentos resultantes en idioma inglés se descargaron en una carpeta con el registro completo en texto sin formato; dichos registros se cargaron a VOSViewer

indicando el tipo de formato de acuerdo con la base de datos de procedencia. Los registros se separaron en tres periodos de años (1972-1989; 1990-2004; y 2005-2017). Para cada periodo se realizó la selección de los parámetros que solicita VOSviewer para realizar el análisis de minería de textos, los cuales dependen del número y tamaño de los archivos que se someten a análisis, y de la profundidad que se desee analizar y visualizar. Es decir, el número mínimo de veces que se repiten las palabras o frases identificadas por VOSViewer en el conjunto de registros; para el periodo 1972-1989, el número mínimo de repetición de las palabras o frases fue de dos; para el periodo 1990-2004, el número mínimo fue cinco, y para el periodo 2005-2017 fue de 10. Dentro del título o resumen de cada registro solo se tomó en cuenta una sola vez cada palabra. Para la normalización de los registros se usó un archivo en texto creado con las palabras o frases a normalizar y a excluir. El método de normalización de las palabras fue la “fortaleza de la asociación”, opción por asignación usada en VOSviewer. Mediante la opción de visualización “overlay visualization” se identificaron los temas más actuales en el rango de años 2007-2015.

RESULTADOS

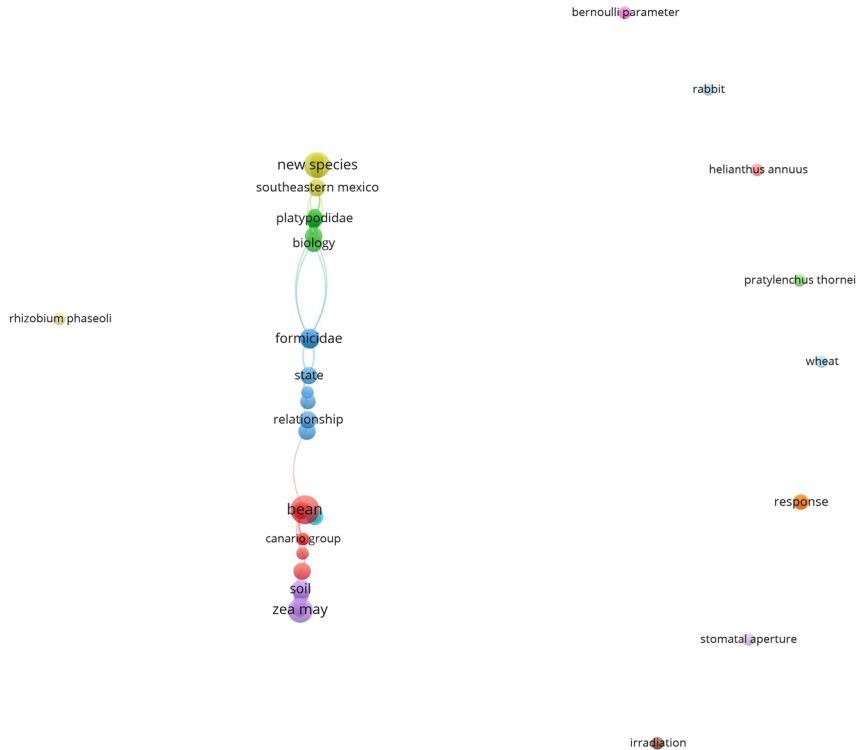
Se identificaron y descargaron 3 457 registros bibliográficos de las bases de datos SCIE y SSCI, y se separaron en periodos similares. De 1972 a 1989 se encontraron 157, que corresponden a 4.54%; del periodo 1990 a 2004 se encontraron 580 registros, que corresponden a 16.78%, y de los años 2005 a 2017 se encontraron 2 720 registros bibliográficos, que corresponden a 78.68% del total.

PERIODO 1972-1989

Se realizó en VOSviewer la minería de textos de los registros bibliográficos del periodo 1972-1989 obtenidos del título y resumen de dichos registros; sólo se tomó en cuenta una palabra o frase

por registro, lo que dio como resultado 537 ítems. El número mínimo de ocurrencias de las palabras o frases en el total de registros fueron dos; lo anterior dio como resultado 58 palabras o frases y se creó el mapa de red temático mostrado en la *Figura 4*.

Figura 4. Mapa de red de las temáticas obtenidas mediante minería de textos de la investigación del Colegio de Postgraduados en revistas de corriente principal periodo 1972-1989.



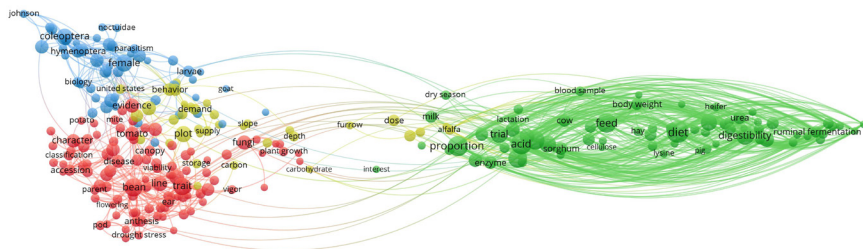
Se identificaron quince grupos temáticos, de los cuales sólo seis están interrelacionados: maíz y suelos, frijol muy relacionado con la acumulación de la fitohormona ácido abscísico, rendimiento, gramíneas y hormigas (*Formicidae*), biología de los escarabajos

ambrosiales, y el estudio de nuevas especies como las Lauráceas, nemátodos, y la *Drosophila mexicana*; también se identificaron nueve temas de investigación que no tienen relación entre ellos: la bacteria *Rhizobium phaseoli*, parámetros de Bernoulli, conejos, girasoles, el nemátodo *Pratylenchus thornei*, trigo, apertura estomatal, irradiación y respuesta *Figura 4*.

PERIODO 1990-2004

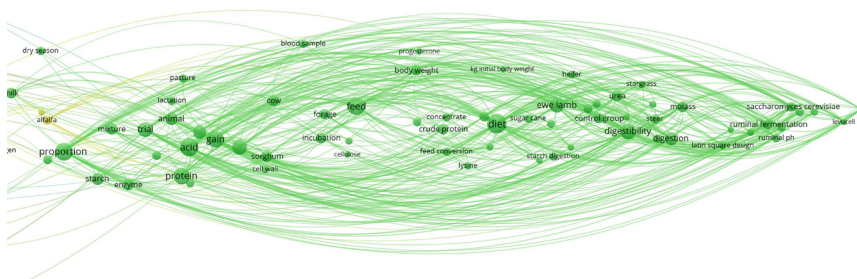
Se realizó minería de textos de los registros bibliográficos del periodo 1990-2004 obtenidos del título y el resumen de dichos registros de los cuales sólo se tomó en cuenta una palabra o frase por registro, lo que dio como resultado 12 472 ítems; el número mínimo de ocurrencias de las palabras o frases en el total de registros fue cinco. Lo anterior dio como resultado 429 palabras o frases, de las cuales se seleccionaron el 60% de los términos más relevantes, lo que resultó en 257 ítems o palabras para realizar el mapa de red temático mostrado en la *Figura 5*. En dicho mapa se identifican cuatro grupos temáticos con dos secciones claramente diferenciadas, por un lado, está un grupo que trata temas de ganadería, y por el otro, tres grupos o *clusters* con temas como producción de cosechas, suelos, y enfermedades y plagas de las plantas.

Figura 5. Mapa de red de las temáticas obtenidas mediante minería de textos de la investigación del Colegio de Postgraduados en revistas de corriente principal periodo 1990-2004.



En el grupo de ganadería, se identificaron temas de investigación como: materia seca, dieta, alimentación, digestión, digestibilidad, fermentación, fermentación ruminal, y forrajes como alfalfa, sorgo, paja de maíz, soya y pastos; otros temas son proteínas, almidón, leche, y levaduras, principalmente *Saccharomyces cerevisiae* y levucell. En cuanto a la investigación sobre animales criados para alimentación humana se destaca la investigación sobre ovinos, bovinos y cerdos. Otros temas destacados son: melaza, suplementación, progesterona, enzimas, lisina, microorganismos y bacterias *Figura 6*.

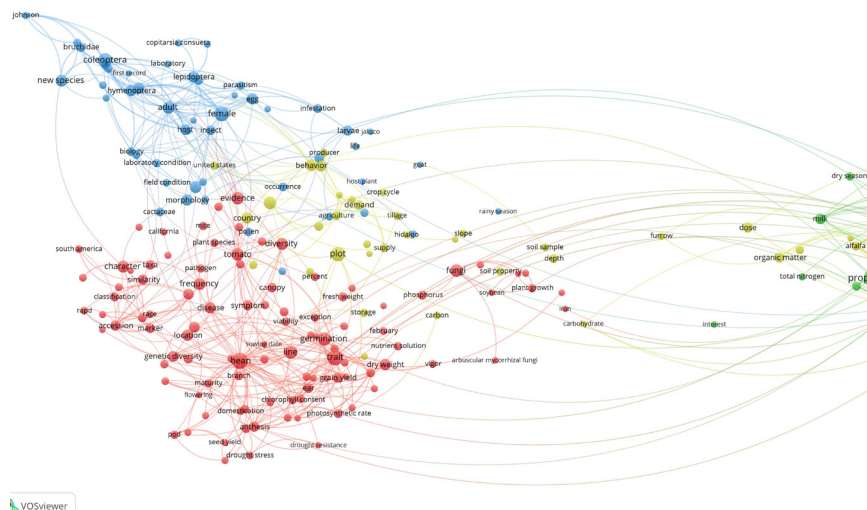
Figura 6. Mapa de red de las temáticas sobre ganadería, obtenidas mediante minería de textos de la investigación del Colegio de Postgraduados en revistas de corriente principal periodo 1990-2004.



En el bloque de tres grupos o *clusters*, se destaca la investigación sobre floración, domesticación, deficiencia y estrés a la sequía en frijol; y la condición y el cultivo de tomate en invernadero. También destacan las investigaciones sobre soya, papa, diversidad genética, germinación de semillas, hongos, hongos micorrícicos y fotosíntesis. Aparecen también en este *cluster* las investigaciones sobre enfermedades de las plantas, ácaros, nemátodos como *Nacobus aberrans*, diversas plagas como *Copitarsia consueta*, insectos como coleópteros (principalmente Brúquidos), himenópteros y lepidópteros. La morfología de cactáceas también es un tema destacado. Otros temas importantes son el muestreo y las propiedades de suelos, materia orgánica, labranza convencional, irrigación,

mercadeo de productos agrícolas, y elementos de importancia para la nutrición vegetal como el carbono, hierro y fósforo (*Figura 7*).

Figura 7. Mapa de red de las temáticas sobre producción de cosechas, suelos y plagas y enfermedades, obtenidas mediante minería de textos de la investigación del Colegio de Postgraduados en revistas de corriente principal periodo 1990-2004.

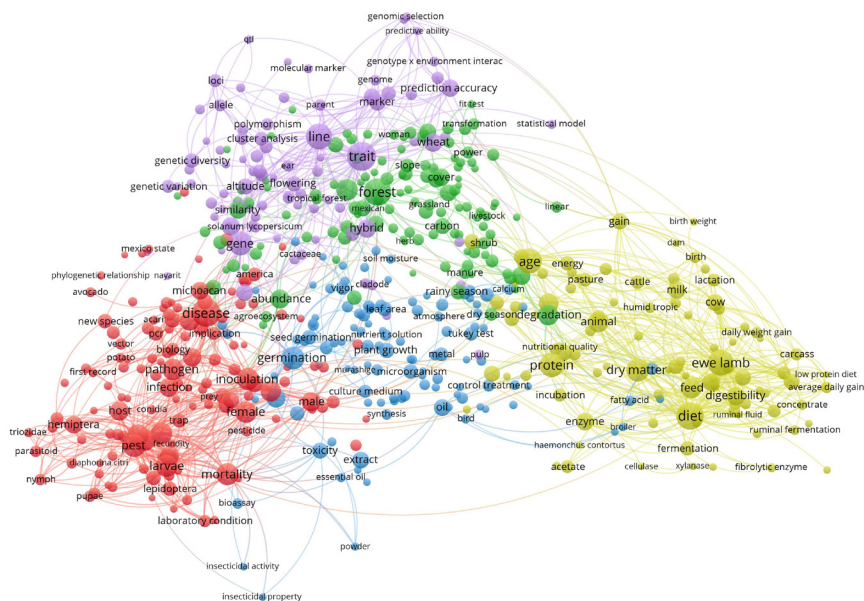


PERIODO 2005-2017

Del proceso de minería de textos aplicada al título y resumen de los documentos, indizados de 2005 a 2017, se obtuvieron 50 934 palabras o ítems, seleccionando las que tenían al menos 10 ocurrencias; resultaron 948 ítems, de las cuales se seleccionaron 60% de las que tienen mayor significancia. Lo anterior dio como resultado un mapa con 569 palabras o frases, con las que se formaron cinco *clusters* o grupos temáticos donde se aprecia un mapa con forma de triple hélice, donde el aspa izquierda contiene los temas relacionados con “producción animal” (color verde olivo), en

el aspa derecha (color rojo) se encuentra el grupo que trata sobre enfermedades y plagas de las plantas, en el aspa superior (color morado) se encuentra el *cluster* que trata sobre genética vegetal, y en medio de las hélices se encuentra dos grupos o *clusters* que tratan sobre suelos y bosques (color verde) y ciencias de las plantas (color azul) (Figura 8).

Figura 8. Mapa de red de las temáticas obtenidas mediante minería de textos de la investigación del Colegio de Postgraduados en revistas de corriente principal periodo 2005-2017.



En el grupo sobre producción animal (color verde olivo), se destaca la investigación sobre dietas, alimentación, suplementación, digestibilidad y desempeño del crecimiento en ovinos; dieta, suplementación y producción de leche en ganado vacuno; dieta, desempeño del crecimiento, y tamaño de la canal en cerdos, y la investigación en caprinos. De manera general, se destaca en este *cluster* la inves-

tigación sobre dieta, ingestión de proteína, digestión, materia seca, la fermentación, degradación ruminal, ganancia de peso, la percepción de la investigación en pastizales y sorgo, y el uso de enzimas para la fermentación de la materia seca (*Figura 8*).

En el *cluster* sobre plagas y enfermedades de las plantas (color rojo), se destaca la investigación insectos y otros organismos fitopatógenos, como ácaros y virus que afectan la producción de cultivos para la alimentación humana y animal; entre las plagas más importantes se destacan: el picudo del agave (*Scyphophorus acupunctatus*), el psílido asiático de los cítricos (*Diaphorina citri*) y el psílido de la papa y tomate (*Bactericera cockerelli*); se destacan también en este grupo los temas, sobre control biológico y control tradicional, así como el uso de los hongos entomopatógenos *Metarhizium anisopliae* y *Beauveria bassiana*. No menos importante se aprecia la investigación sobre infecciones y necrosis, hongos micorrízicos, biofertilizantes y nematodos, principalmente *Nacobbus aberrans*; también se destaca la investigación de enfermedades sobre chile (*Capsicum annuum*) como *Phytophthora capsici*, y en otros cultivos como aguacate, guayaba, papa, papaya, plátano y mango y sorgo (*Figura 8*).

En los temas de investigación sobre genética vegetal (color morado), se destacan aquellos sobre características, genes y líneas de diferentes cultivos, híbridos (floración y llenado de grano) caracterización morfológica, mejoramiento genético, diversidad genética, variabilidad genética, selección de plantas, selección genómica, marcadores moleculares, polimorfismo, exactitud de la predicción, granos (principalmente trigo, maíz, haba y sorgo), y producción de semillas; en este *cluster* aparece también la investigación sobre cactáceas (*Figura 8*).

El *cluster* sobre cultivos agrícolas y bosques (color verde), muestra temas de investigación sobre bosques y su relación con el carbono en el suelo, producción maderera, especies y abundancia de árboles; se agrupan además en este *cluster* los temas sobre sistemas agrícolas, fertilidad de suelos, degradación de materia orgánica, pastizales, medio ambiente, y ecosistemas. Llama la atención que el tema sobre género se agrupa en este *cluster* (*Figura 8*).

El grupo de investigación en ciencia de las plantas (color azul) trata sobre diversos aspectos como el crecimiento de las plantas, la propagación, emergencia y la germinación de las semillas. Destacan, además, los temas sobre déficit de agua y salinidad; los minerales: potasio, fósforo, calcio, hierro y cobre; la actividad antioxidante, flavonoides, metabolitos, taninos y compuestos fenólicos; se agruparon aquí, además, temas como micorrizas, soluciones nutritivas, cultivo de tejidos, medios de cultivo, fitoremediación, extractos de plantas, aceites, actividad y propiedades insecticidas de plantas, conductividad eléctrica, prueba de Tukey, y aguas negras (*Figura 8*).

Los temas de investigación más actuales identificados (color amarillo en la *Figura 9* en los diferentes *clusters* mediante el análisis realizado son: agentes de control biológico, triózidos (*Trioziidae*), psílido asiático de los cítricos (*Diaphorina citri*), psílido de la papa (*Bactericera cockreli*), factores abióticos, chile, tomate, rendimiento y peso de frutos, alto rendimiento, híbridos comerciales, plantas medicinales, selección genómica, interacción genotipo ambiente, aceites esenciales, capacidad antioxidante, exactitud de la predicción, el parásito del estómago de rumiantes *Haemonchus contortus*, canal (de animales), prueba de Tukey y tratamiento de aguas residuales.

En el periodo de 1990 a 2004, la investigación empezó a tomar

documentos, que corresponden a 16.78% del total. Con tales registros, se formó un mapa donde se identificaron dos secciones con cuatro grupos o *clusters*; la primera sección está formada por el *cluster* de ganadería, y la restante sección está conformada por tres grupos o *clusters* con temas como producción de cosechas, suelos, y enfermedades y plagas de las plantas.

En el último periodo analizado, de 2005 a 2017, la investigación creció enormemente, y se llegaron a publicar en promedio casi 210 documentos por año, hecho que contrasta enormemente con el periodo 1972-1989, cuando sólo se publicaron ocho y medio documentos por año. Con estos registros, se formó un mapa con cinco *clusters* o grupos temáticos: ganadería, enfermedades y plagas de las plantas, genética vegetal, suelos y bosques, y ciencias de las plantas.

Entre los temas de investigación más actuales, se encuentran agentes de control biológico, triózidos y psílidos; algunos cultivos de gran consumo en México como, chile y tomate; híbridos comerciales, plantas medicinales, selección genómica, aceites esenciales, capacidad antioxidante, el parásito del estómago de rumiantes *Haemonchus contortus*, y el tratamiento de aguas residuales.

La minería de textos es una metodología que nos ayuda a encontrar información inmersa en los títulos y resúmenes de documentos como artículos científicos, que no están a simple vista. Esto abre una vía de investigación que ayuda a identificar las temáticas de investigación en documentos científicos; sin embargo, debe ser tratada con cuidado ya que no es una metodología sobre la que se tenga control del análisis de los documentos mediante los términos o palabras dentro de los registros, como sí sucede con las palabras clave o descriptores.

Se recomienda que, para tener una visión más completa de las temáticas de investigación del Colegio de Postgraduados, se realice un análisis de palabras conjuntas o co-palabras con las palabras clave de los registros bibliográficos proporcionadas por los autores de los documentos y por los indizadores de las bases de datos SCIE y SSCI.

BIBLIOGRAFÍA

- Centre for Science and Technology Studies. «VOSviewer - Visualizing Scientific Landscapes». VOSviewer, 2018. <http://www.vosviewer.com//>.
- Colegio de Postgraduados. «Colegio de Postgraduados». Conócenos, 2016. <http://www.colpos.mx/wb/index.php/conocenos/>.
- . «Línea de Tiempo». Conócenos, 2014. <http://www.colpos.mx/wb/index.php/conocenos/linea-de-tiempo>.
- . «Líneas de Generación y/o Aplicación del Conocimiento Institucionales». Investigación, 2018. <http://www.colpos.mx/wb/index.php/investigacion/lineas-de-generacion-y-o-aplicacion-del-conocimiento-institucionales>.
- . «Oferta Educativa». Educación, 2017. <http://www.colpos.mx/wb/index.php/educacion/oferta-educativa>.
- Eck, Nees Jan Van, y Ludo Waltman. «Text mining and visualization using VOSviewer». *arXiv:1109.2058 [cs]*, 9 de septiembre de 2011. <http://arxiv.org/abs/1109.2058>.
- . «VOS: A New Method for Visualizing Similarities Between Objects». En *Advances in Data Analysis*, editado por Reinhold Decker y Hans -J. Lenz, 299-306. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg, 2007.
- Feldman, Ronen, y James Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: Cambridge University Press, 2006. <https://doi.org/10.1017/CBO9780511546914>.
- Gobster, Paul H. «(Text) Mining the LANDscape: Themes and trends over 40 years of Landscape and Urban Planning». *Landscape and Urban Planning* 126 (1 de junio de 2014): 21-30. <https://doi.org/10.1016/j.landurbplan.2014.02.025>.
- González Cossío, Félix. «Prólogo». En *Nuevas tendencias científicas y tecnológicas en el Colegio de Postgraduados*, 5-7. Montecillo, Texcoco, Estado de México: Colegio de Postgraduados, 2010.

- Larqué-Saavedra, Alfonso. *Crónicas de la ciencia 2005-2014*. Mérida, Yucatán: CICY, Consejo Consultivo de Ciencias de la Presidencia de la Republica, 2014.
- Salager-Meyer, Françoise. «Peripheral Scholarly Journals: From Locality to Globality». *Ibérica* 30 (1 de noviembre de 2015): 15-36.
- Tan, Ah-Hwee. «Text Mining: The State of the Art and the Challenges». En *In Proceedings PAKDD'99 Workshop on Knowledge Discovery from Advanced Databases (KDAD'99)*, 71-76, 1999. http://www.ntu.edu.sg/home/asahtan/papers/tm_pakdd99.pdf.

Manejo de datos. Una aproximación desde los estudios de la información. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Francisco Xavier González y Ortiz; revisión de pruebas, Valeria Guzmán González; formación editorial, Natalia Cristel Gómez Cabral. Instituto de Investigaciones Bibliotecológicas y de la Información / UNAM. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en febrero de 2020.