

LA MINERÍA DE DATOS COMO HERRAMIENTA DE ANÁLISIS EN LA CONSERVACIÓN DE RECURSOS FITOGENÉTICOS

Línea temática: Informática y Sistemas de Información

Karina G. Fernández Hidalgo¹

¹Departamento de Informática. Facultad de Ciencias Exactas, Físicas y Naturales.
Universidad Nacional de San Juan. Avda. Ignacio de la Roza y Meglioli. Rivadavia.

karinaferh@gmail.com

RESUMEN

El procesamiento de información basada en minería de datos está orientada al análisis de atributos provenientes de un centro de conservación de recursos fitogenéticos (RFN) de la Provincia de San Juan, entiéndase como RFN a los materiales provenientes del reino vegetal que poseen un valor real o potencial para alimentación y/o para la agricultura. Este centro de conservación de la provincia,

posee registrados en planillas planas y en papel, numerosos datos biológicos referidos a especies vegetales de la flora nativa y también de especies y variedades de cultivo tradicional en la zona. Se propone sistematizar, segmentar y clasificar esos datos biológicos para obtener agrupaciones requeridas para la planificación, manejo, evaluación, regeneración y transferencia del material (semillas) conservado.

Las tareas principales radicarán en el uso de algoritmos específicos que permitan tareas de segmentación y clasificación de las colecciones de semillas de los dos grupos de especies, caratuladas mediante tablas con más de 40 atributos cada una y con diferente tipología de datos asociada. Se hará uso de la herramienta de software RapidMiner (versión 5.3.012), entorno de aplicación de algoritmos de aprendizaje de máquina y minería de datos, con accesibilidad a la mayoría de los gestores de bases de datos. Se aplicarán todos los pasos involucrados en la minería de datos desde el pre-procesamiento hasta la visualización de resultados al evaluar diferentes estrategias de segmentación, de clasificación y de reglas de asociación mediante una interfaz amigable y que se ofrece bajo una licencia AGPL versión 3.0 (software libre).

Palabras claves: minería de datos, Rapid Miner, Recursos Fitogenéticos

ANTECEDENTES

La actividad de conservación *ex situ* de recursos fitogenéticos en la provincia de San Juan, está siendo fortalecida mediante la ejecución del Proyecto “Desarrollo competitivo del sector semillero de la provincia de San Juan”, el cual es financiado por el Banco Mundial y con aportes locales en el marco del Ministerio de Producción y Desarrollo Económico del Gobierno de la Provincia, los cuales son administrados por el Programa de Servicios Agrícolas Provinciales (PROSAP/BIRF 7597-AR). Este proyecto se lleva a cabo con centro en el Instituto de Investigación y Desarrollo Agroindustrial Hortícola Semillero (INSEMI).

Por otro lado, en el ámbito de la Universidad Nacional de San Juan (UNSJ), se encuentra en ejecución el Proyecto “Minería de Datos en la Determinación de Patrones de Uso y Perfiles de Usuarios (21/E889)”, financiado por CICTICA-UNSJ trienio 2011-2013, con el cual y atendiendo a la temática abordada, existe la posibilidad de un trabajo común de fuerte integración. En este contexto se cuenta con el aporte de la beca de investigación CICTICA otorgada a la Alumna Karina Fernández, periodo 2013/2014 para el estudio de minería de datos en la segmentación y clasificación de un banco de germoplasma.

En el corto plazo se contará además con aportes del Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCyT), para el fortalecimiento de las bases de datos y para la formación de recursos humanos, ya que el INSEMI ha sido adherido al Sistema Nacional de Datos Biológicos (SNDB – Res 05/12)

FUNDAMENTACIÓN

En los últimos 60 años aproximadamente, el hombre comprendió la importancia que reviste conservar la variabilidad genética, ya que depende de ella en una relación directa. Por ello se idearon e implementaron distintos planes de conservación, entendiéndose por ello “la ordenación para el empleo humano de los recursos genéticos para que éstos puedan remitir los máximos beneficios sostenibles para las generaciones actuales, manteniendo aun su valor potencial para satisfacer las necesidades y aspiraciones de las generaciones futuras” (UICN, 1980).

Se entiende por Recursos Genéticos, la variabilidad genética acumulada en todos los organismos vivos a lo largo de millones de años de evolución (REGENSUR, 2007). Entonces los Recursos Fitogenéticos se definen como cualquier material genético de origen vegetal de valor real o potencial para la alimentación y la agricultura, y que constituyen la materia prima utilizada por los agricultores, por consiguiente resulta fundamental su conservación y el mantenimiento para una producción agrícola sostenible (FAO, 1996; UICN, 1980).

Con la finalidad de conservar los recursos genéticos del país, nacen los Bancos de Germoplasma, con el objeto de preservar las características de material vegetal que forman parte del patrimonio argentino. Los bancos aspiran a proteger especies de interés que satisfacen una demanda actual, así como para aquellas que aún no presentan características de uso inmediato y podrían ser consideradas valiosas en el futuro.

Por otra parte, representan la salvaguardia de especies en peligro de extinción de aquellas zonas que si bien existe en la actualidad una enorme cantidad y variedad de recursos de información relacionados con la biodiversidad, su aprovechamiento eficiente depende de la facilidad para encontrarlos, indagar sus contenidos y determinar sus alcances y limitaciones (Rivera-Gutiérrez et al., 2003).

La documentación de la información recopilada y generada es fundamental para la toma de decisiones sobre el manejo de las colecciones en un banco de germoplasma, y sobre su valor de uso (Engels et al., 2007). Por otro lado, el valor del germoplasma aumenta en la medida que se conoce, de ahí la importancia de una adecuada documentación de los datos recopilados y generados (Rao et al., 2007).

El sistema de documentación se utiliza para la recuperación, el almacenamiento, el mantenimiento o la actualización, procesamiento y análisis, e intercambio de datos con otros centros de conservación (Rivera-Gutiérrez et al., 2003).

El éxito de la revolución digital y el crecimiento de Internet aseguran que grandes volúmenes de datos de alta dimensión, están disponibles en todo lo que nos rodea. Esta información se mezcla a menudo, con la participación de diferentes tipos de datos tales como texto, imagen, audio, voz, hipertexto, gráficos y componentes de vídeo entremezcladas unas con otras. Sin embargo, a menudo la mayor parte de estos datos no son de mucho interés para la mayoría de los usuarios.

El Descubrimiento de Conocimiento en Bases de Datos (KDD) es un análisis automático y exploratorio que permite el modelado de depósitos de datos de gran

tamaño. KDD es el proceso organizado de identificar patrones válidos, novedosos, útiles y comprensibles a partir de extensos y complejos conjuntos. Data Mining (DM) es el núcleo del proceso de KDD, que implica la inferencia de los algoritmos que analizan y modelan los datos, y permite descubrir patrones previamente desconocidos y de interés para los usuarios. Se trata de un área de investigación y desarrollo interdisciplinaria que abarca diversos dominios, y lejos de estar saturada, se amplía con nuevas técnicas y orientaciones (Mucherino A., y otros. 2009; O. Maimon O. Rokach L. 2010; Witten, Frank, Hall, 2011).

El procesamiento de información basada en minería de datos, está orientada a análisis de atributos pertenecientes al área de conservación ex situ de recursos fitogenéticos de San Juan. Se mantienen colecciones de herbario y de semillas tanto de especies nativas como también de cultivos tradicionales, importantes para la provincia. Este material genético responsable de las características de una planta que se transmite de una generación a la siguiente, es para el futuro beneficio de la humanidad y del ambiente.

En este contexto, contar con datos digitalizados y una base de datos actualizada, ágil y de fácil acceso, que contenga información georreferenciada y vinculante sobre los recursos fitogenéticos de San Juan, resulta además, una herramienta de alto valor, dado que no se han encontrado aún en la zona, instituciones que generen, mantengan y actualicen en forma continua datos biológicos.

Se plantea aplicar minería de datos a los diferentes registros colectados en el área de conservación de recursos fitogenéticos de San Juan. Las tareas principales radicarán en el uso de algoritmos específicos que permitan tareas de segmentación y clasificación de datos provenientes de las colecciones de semillas de especies correspondientes a la geografía de San Juan. Para ello se hará uso de la herramienta de software RapidMiner cuya última versión disponible es la 5.3.012. Se trata de un entorno de aplicación de algoritmos de aprendizaje de máquina y minería de datos, fácil de instalar y de ejecutar en cualquier plataforma y sistema operativo. Allí se pueden aplicar todos los pasos involucrados en la minería de datos desde el pre-procesamiento hasta la visualización de resultados al evaluar diferentes estrategias de

segmentación, clasificación y reglas de asociación mediante una interfaz amigable (North 2012) y que se ofrece bajo una licencia AGPL versión 3.0 (software libre).

Por otro lado, se hará un análisis exhaustivo de aquellas herramientas de captura y gestión de colecciones biológicas (HCGC) propuestas por el banco de Germoplasma como son Biótica, Specify, Herbar, con el fin de centralizar, publicar y acceder a los datos de registros biológicos sobre Internet, para unificar las bases de datos con los requerimientos que exige de la Secretaría de Agricultura.

Objetivo general

Segmentar y clasificar datos biológicos provenientes de Recursos Fitogenéticos de San Juan para obtener agrupaciones requeridas para la planificación, manejo, regeneración y transferencia del germoplasma conservado.

Objetivos específicos

- Estudiar introductoriamente el área de conocimiento e Interpretar los datos.
- Relevar y analizar las herramientas de software propuestas por el Banco de germoplasma del Instituto de Investigación y desarrollo hortícola semillero (INSEMI) dependiente de la Secretaría de Agricultura y necesidades de hardware asociado.
- Extraer conocimiento relevante de los datos que favorezca la toma de decisiones para el manejo de las colecciones conservadas mediante la aplicación de herramientas específicas de minería de datos.
- Evaluar alternativas de visualización de resultados que permita una óptima toma de decisiones con respecto a la planificación de las campañas de recolección de recursos fitogenéticos.

DESARROLLO

Los datos sobre los recursos fitogenéticos conservados se encontraban 100% almacenados en planillas de recolección y cuadernos de campo, por lo que la primera

tarea y la que más tiempo insumió se corresponde con la digitalización de los mismos, además de la limpieza, integración, transformación y reducción de los datos (pre-procesamiento) que tiene por objetivo preparar los datos para permitir posteriormente la migración de los mismos a una base de datos adecuada y poder realizar los análisis propuestos.

La importancia de dicha tarea de pre-procesamiento es por la calidad de los datos sobre los que se aplican técnicas de descubrimiento de conocimiento ya que esto impacta de manera directa en la calidad del conocimiento que se descubre a partir de ellos. (Cruz Arrela, 2010.)

The image shows two versions of a data collection form for native species germplasm. The left form is the main collection form, and the right form is a detailed version for the sampled specimen. Both forms are from the 'Banco de Germoplasma de San Juan' and include sections for general collection data, species identification, collection location, and detailed morphological characteristics of the specimen.

Fig 1: Constructo relevado en campo de especie Nativa

En la Fig 1 se puede apreciar el modelo de la planilla de papel de semillas nativas con las que el equipo colector cuenta al momento de la recolección y en la que anota toda la información requerida.

Hasta la fecha se ha logrado un elevado porcentaje de la sistematización y posterior digitalización de la información correspondiente a las colecciones de semillas de la flora nativa biológica, como así también de las colecciones de semillas de los cultivos tradicionales. Estos registros son caracterizados por aproximadamente 80 atributos multivaluados. Estos datos han sido categorizados en planillas Excel, a partir de las cuales serán trasladados a la base de datos que se establecerá para el área

de conservación de recursos fitogenéticos del INSEMI y desde donde la herramienta de aplicación RM tomará los datos, previa compatibilización en los tipos de datos asociados a los atributos, que permita a su vez, viabilizar la aplicación de diferentes algoritmos de segmentación y clasificación. Esto último está iniciándose y los resultados que se obtengan de estos análisis definirán los próximos pasos a seguir dentro del ámbito de la minería de datos.

Row No.	Elemento	Cuaderno	Página	Fecha recolección	Responsable	Otros colectores	Colección	Especie	Nombre vulgar	Familia	Ejemplar s.	Tipo de mu.
1	1	?	?	22/06/2011 0:00:00 ART	Juan Scaglia	Cristian Piedra	CAESALPINEACEAE	Caesalpinia paraguayensis	Cesáspinea	CAESALPINOIDEAE	Scaglia, 20	fruto
2	2	?	?	27/02/2011 0:00:00 ART	Martin Hadad	Juan Scaglia	CAESALPINEACEAE	Cerodidum praecox (R&P) H	Brea	CAESALPINACEAE	Hadad, 47	fruto
3	3	?	?	30/12/2011 0:00:00 ART	Juan Scaglia	?	CAESALPINEACEAE	Zuccagnia punctata Cav.	Jarilla macho	CAESALPINACEAE	Scaglia, 39	fruto
4	4	4	22	19/01/2012 0:00:00 ART	Juan Scaglia	?	CAESALPINEACEAE	Cerodidum praecox subsp g	Brea	CAESALPINACEAE	Scaglia, 47	fruto
5	5	1	34	04/01/2011 0:00:00 ART	Martin Hadad	Juan Scaglia	CAESALPINEACEAE	Cerodidum praecox	Brea	CAESALPINACEAE	Hadad, 27	fruto
6	6	?	?	20/01/2012 0:00:00 ART	Juan Scaglia	Gerardo de la Vega	CAESALPINEACEAE	Cerodidum praecox var. Prae	Brea	CAESALPINACEAE	Scaglia, 44	fruto
7	1	?	?	25/05/2010 0:00:00 ART	Martin Hadad	Juan Scaglia	PAPILIONOIDEACEAE	Ramoniae grolae	Chica	PAPILIONOIDEAE	Hadad, 25	fruto
8	2	?	?	21/04/2011 0:00:00 ART	Juan Scaglia	?	PAPILIONOIDEACEAE	Geoffroea decorticans	Chañar	PAPILIONOIDEAE	Scaglia, 9	fruto
9	1	1	128	21/04/2010 0:00:00 ART	Martin Hadad	Uta Karlin	MMOSACEAE	Prosopis alata Phil.	Alpataco, Ismar	MMOSACEAE	Hadad, 14	fruto
10	2	?	?	18/11/2010 0:00:00 ART	Juan Scaglia	?	MMOSACEAE	Mimozganthus carinatus (Gf	Lata	MMOSACEAE	Scaglia, 34	fruto
11	3	1	1	21/03/2010 0:00:00 ART	Martin Hadad	?	MMOSACEAE	Acacia atramentaria	Espirillo	MMOSACEAE	Hadad, 5	fruto
12	4	3	60	16/03/2011 0:00:00 ART	Juan Scaglia	Mauricio Perez	MMOSACEAE	Acacia visco	Visco	MMOSACEAE	Scaglia, 60	fruto
13	5	?	?	27/05/2011 0:00:00 ART	Juan Scaglia	Rodolfo Bricuela, José C	MMOSACEAE	Acacia atramentaria	Espirillo negro	MMOSACEAE	Scaglia, 14	fruto
14	6	1	4	24/03/2010 0:00:00 ART	Martin Hadad	Yarina Ribas	MMOSACEAE	Acacia atramentaria Benth.	Aromo negro	MMOSACEAE	Hadad, 7	fruto
15	7	4	18	18/01/2013 0:00:00 ART	Juan Scaglia	Rodolfo Bricuela, Carlos	MMOSACEAE	Prosopis flexuosa	Algarobo negro	MMOSACEAE	Scaglia, 80	fruto
16	8	4	16	13/01/2013 0:00:00 ART	Juan Scaglia	Rodolfo Bricuela	MMOSACEAE	Prosopis flexuosa	Algarobo negro	MMOSACEAE	Scaglia, 77	fruto
17	9	4	15	14/01/2013 0:00:00 ART	Juan Scaglia	Rodolfo Bricuela, Javier J	MMOSACEAE	Prosopis chilensis	Algarobo blanco	MMOSACEAE	Scaglia, 78	fruto
18	10	1	6	22/04/2010 0:00:00 ART	Martin Hadad	Laura Aguiar, Natalia An	MMOSACEAE	Acacia furcata spina	Garabato	FABACEAE	Hadad, 17	fruto
19	11	1	7	24/05/2010 0:00:00 ART	Martin Hadad	Juan Scaglia	MMOSACEAE	Acacia furcata spina	Garabato	MMOSACEAE	Hadad, 24	fruto
20	12	1	10	09/05/2010 0:00:00 ART	Martin Hadad	?	MMOSACEAE	Acacia visco	Visco	MMOSACEAE	Hadad, 20	fruto
21	13	1	8	02/05/2010 0:00:00 ART	Martin Hadad	?	MMOSACEAE	Acacia furcata spina	Garabato	MMOSACEAE	Hadad	fruto
22	14	1	?	04/04/2010 0:00:00 ART	Martin Hadad	Juan Scaglia	MMOSACEAE	Acacia aroma Giles ex Hoo	Aromo	MMOSACEAE	Hadad, 11	fruto
23	15	?	?	01/06/2011 0:00:00 ART	Juan Scaglia	?	MMOSACEAE	Acacia caven	Espirillo	MMOSACEAE	?	fruto

Fig 2: Registros de especie nativa en la herramienta RM (Vista Data View)

En la Fig 2 se puede apreciar la lectura realizada desde RM de una porción de la fuente de datos original excel. En esta primer tanda de registros cargados se observa la existencia de 80 atributos entre los que figuran datos asociados al personal de la campaña de recolección, lugar físico de ubicación de los datos originales, que si bien pueden ser de suma importancia para la determinación de responsabilidades en el INSEMI puede tener escasa importancia a la hora comenzar con las tareas de segmentación y clasificación. Por lo expuesto anteriormente, el grupo de trabajo está en plena tarea de definición de atributos relevantes inherentes a la etapa de preprocesamiento de DM.

Específicamente, en cuanto a la tarea de preprocesamiento se verá un ejemplo:

Al preprocesar la planilla de semillas nativas, examinamos y observamos que existen datos que significan lo mismo pero difieren en mayúsculas, minúsculas y acento. Por

lo que fue necesario realizar un análisis exhaustivo de cada atributo con sus respectivos valores.

Role	Name	Type	Statistics	Range	Missings
regular	Fecha colección	date_time	length = 1041 days	[14-mar-2010 0:00:00 ART ; 18-ene-2012	0
regular	Responsable de campaña	binominal	mode = Juan Scaglia (53), least = Juan Scaglia (53), Martin Hadad (29)		1
regular	Colección	polynominal	mode = MIMOSACEAE (22), least = CAESALPINEACEAE (6), PAPILIONOIDE		0
regular	Especie	polynominal	mode = Prosopis chilensis (6), le = Prosopis chilensis (6), Bulnesia retama		0
regular	Nombre vulgar	polynominal	mode = Retamo (7), least = Cesa = Cesalpineae (1), Brea (4), Jarilla macho		2
regular	Familia	polynominal	mode = MIMOSACEAE (22), least = CAESALPINOIDEACEAE (1), CAESALPIN		0
regular	Ejemplar de herbario	polynominal	mode = Scaglia, 20 (1), least = Sr = Díaz Bisutti, 36 (1), Hadad, (1), Hadad, 1		4
regular	Tipo de muestra colectada	polynominal	mode = fruto (77), least = semilla = fruto (77), fruto y semilla (4), semilla (2)		0
regular	Cantidad aprox frutosemillas colectadas	integer	avg = 1730.120 +/- 2633.016	[30.000 ; 15000.000]	0
regular	Número aprox de frutosemillas colectadas	integer	avg = 4071.852 +/- 3552.612	[70.000 ; 20000.000]	2
regular	Material vegetativo (ramas)	binominal	mode = No (83), least = No (83)		0
regular	Condición de la muestra	polynominal	mode = Nativa (81), least = Arbol = Arbolado publico (1), Nativa (81), Natural		0
regular	Frecuencia	polynominal	mode = Frecuente (46), least = Ej = Ejemplar unico (1), Muy frecuente (20), F		0
regular	Provincia	binominal	mode = San Juan (83), least = Sa = San Juan (83)		0
regular	Departamento	polynominal	mode = Valle Fertil (17), least = Sr = Albaridon (2), Iglesias (5), Ullum (6), Cau		1
regular	Localidad	polynominal	mode = Ullum (5), least = La = Albaridon (2), Iglesias (5), Ullum (6), Caucete (13), VALLE FERTIL (2), Jachal (12), valle fertil (1), Rivadavia (4), San Martin (1), Valle Fertil (17), Zonda (2), Valle Fértil (4), Calingasta (3), Sarmiento (6), 25 de Mayo (4)		0
regular	Altitud (msnm)	integer	avg = 925.595 +/- 438.506		0
regular	Pendiente del terreno	numeric	avg = 0.060 +/- 0.060		0
regular	Exposición	polynominal	mode = Este (29), least = Sur = Sur (29)		0
regular	Latitud	integer	avg = 30.620 +/- -0.606		0
regular	Lat-Minut	real	avg = 28.910 +/- -17.397		0
regular	Longitud	integer	avg = 67.696 +/- -0.627	[67.000 ; 69.000]	4
regular	Long-Minut	real	avg = 26.837 +/- -14.861	[0.876 ; 58.397]	4

Fig 3: Registros de especie nativa en la herramienta RM (Vista Meta Data View)

En la Fig 3 tomada de una porción de la fuente de datos original excel, se puede apreciar el atributo Departamento, con sus variantes; entre ellas VALLE FERTIL (2 registros), valle fertil (1 registro), Valle Fertil (17 registros) y Valle Fértil (4 registros), interpretadas de manera distinta solo por diferir de un acento, mayúscula o minúscula, por lo que se hace necesario unificar todos los datos a Valle Fertil.

Otro ejemplo interesante se presenta en las planillas cultivadas:

En el atributo Productor= “Celso Paredes”; en distintos registros del atributo “Importancia de la conservación de semillas”, respondió lo siguiente:

- Para seguir sembrando en el futuro
- La parece bien para poder seguir sembrando mas adelante
- Le parece bien, además de seguir trabajando la siembra
- Le parece bien, además de que se siga sembrando mas adelante
- Le parece bien para que se siga trabajando en la siembra mas adelante
- Le parece bien para que se siga trabajando en la siembra mas adelante
- Le parece bien, además de que se siga sembrando mas adelante

Lo cual es necesario unificar las respuestas pues todas significan lo mismo.

Una vez que se realiza el proceso anterior para cada productor, se vuelve a analizar la totalidad de los registros y los resultados son los siguientes:

- 22 registros Para que no se pierdan
- 41 registros Le parece bien para que no se pierdan
- 1 registro Le parece bien para que no se terminen de perder
- 2 registros Porque hay muchas semillas que están desapareciendo
- 1 registro Es importante para que no se pierda la semilla

La palabra clave es: pierdan, desapareciendo, perder. Todos los registros están dando el mismo mensaje, que es que “las semillas no se pierdan”. Por lo tanto se unifican todas las respuestas seleccionando la que tiene mayor frecuencia de repetición.

En consiguiente a la tarea de preprocesamiento se proseguirá con el relevamiento y el análisis de las herramientas de software y necesidades de hardware asociado, y con la generación de conocimiento relevante de los datos que favorezca la toma de decisiones para el manejo de las colecciones conservadas mediante la aplicación de herramientas específicas de minería de datos.

CONSIDERACIONES FINALES

El presente trabajo es un avance dentro de la beca de investigación “Extracción de conocimiento en un Banco de Germoplasma”. La información recopilada hasta el momento corresponde a características genéticas de la fitogeografía sanjuanina, en concreto especies de la flora nativa de la provincia y a especies de cultivo tradicional. Cada muestra de semilla conservada posee datos descriptivos de la especie y del sitio de colección. En el caso de los cultivos tradicionales, se cuenta además con información sobre el productor donante, la edad del cultivo y la forma de obtención de las semillas (herencia familiar, compra directa, regalo de vecinos, etc). Y a partir de la tarea de preprocesamiento que insumió tanta cantidad de tiempo, se busca utilizar esa información procesada para identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles con el fin de aportar

conocimiento para la toma de decisiones, aplicando diferentes algoritmos de Segmentación y Clasificación existentes en la herramienta RapidMiner; cabe aclarar que con dichos algoritmos, se han analizado los datos provistos por 83 planillas correspondientes a semillas Nativas y 316 planillas de semillas Cultivadas, dado que el número de planillas resulta insuficiente, la aplicación de las herramientas sólo tiene como objetivo ser un ejemplo ilustrativo.

Se pretende además, y en el marco de un trabajo integrador de titulación intermedia de la carrera licenciatura en Ciencias de la Computación, desarrollar una herramienta de software e interfaz adecuada, que permita la carga directa de la información durante la campaña de recolección, en dispositivos tipo tablets o celular avanzado a cargo del personal afectado, logrando una mayor agilidad al momento de carga y procesamiento de datos

BIBLIOGRAFÍA

- REGENSUR/PROCISUR. 2007. Acceso a los recursos genéticos: estado de situación en los países del cono sur.
- Unión Mundial para la Conservación de la Naturaleza (UICN). 1980. Estrategia Mundial para la Conservación.
- FAO. 1996. Plan de Acción Mundial para la Conservación y la utilización sostenible de los Recursos Fitogenéticos para la Alimentación y la Agricultura y la Declaración de Leipzig. Cuarta Conferencia Técnica Internacional sobre Recursos Fitogenéticos, Leipzig, Alemania, 17-23 de junio, 64 pp.
- Engels, J.M.M. y Visser, L. (eds.). 2007. Guía para el manejo eficaz de un banco de germoplasma. Manuales para Bancos de Germoplasma No. 6. Bioversity International, Roma, Italia. ISBN 978-92-9043-767-3
- Rao, N.K., J. Hanson, M.E. Dulloo, K. Ghosh, D. Novell y M. Larinde. 2007. Manual para el manejo de semillas en bancos de germoplasma. Manuales para Bancos de Germoplasma No. 8. Bioversity International, Roma, Italia. ISBN 978-92-9043-757-4.

-
- Rivera-Gutiérrez H. F., Suárez-Mayorga A. M., Varón-Londoño A. 2003. Estándar para la documentación de metadatos de conjuntos de datos relacionados con biodiversidad, versión (electrónica). Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia, 61 p.
 - Mucherino A., Papajorgji P., Pardalos P. 2009. “Data Mining in Agriculture” (Springer Optimization and Its Applications Volume 34)
 - Witten I.; Frank E., Hall M. 2011 “Data Mining Practical Machine Learning Tools and Techniques”, Third Edition (The Morgan Kaufmann Series in Data Management Systems) Morgan Kaufmann
 - Maimon, O. Rokach L. 2010. “Data Mining and Knowledge Discovery Handbook” –Springer
 - Cruz Arrela, Liliana. Tesis de Grado de la Universidad Nacional Autónoma de México. Marzo 2010.
 - North M. 2012. “Data Mining for the Masses” ISBN: 0615684378. A Global Text Project Book.