

La Estadística Implicativa como gestor del conocimiento en el análisis de la investigación universitaria en la Escuela Superior Politécnica de Chimborazo

**Rubén Pazmiño (1), Francisco Peñalvo (2), Miguel Conde (3),
Paola Chiluiza (4), Diana Campoverde (5)**

(1) CIED - Escuela Superior Politécnica de Chimborazo - Ecuador, rpazmino@esPOCH.edu.ec. (2) Universidad de Salamanca-España, fgarcia@usal.es. (3) Universidad de León - España, miguel.conde@unileon.es. (4) CIED - Escuela Superior Politécnica de Chimborazo - Ecuador, paola.chiluiza@esPOCH.edu.ec. (5) CIED - Escuela Superior Politécnica de Chimborazo - Ecuador, diana.campoverde@esPOCH.edu.ec.

Resumen: El Análisis Estadístico Implicativo permite gestionar el conocimiento descubriendo reglas de cuasi implicación entre variables. Esta investigación utiliza la Estadística Implicativa para determinar si la investigación universitaria concuerda con las líneas y programas de investigación ya definidos en la Escuela Superior Politécnica de Chimborazo. Nos basamos en el análisis de 970 trabajos de titulación realizados desde el año 2012 hasta el año 2016 en la Facultad de Ciencias de la ESPOCH. Concluimos que hay una relación positiva de similitud, cohesión e implicación entre los trabajos de titulación, las líneas de investigación, los programas de investigación y las carreras. Este es un trabajo de investigación replicable y aplicable a otros contextos educativos. Esta investigación aporta con una nueva metodología para la gestión del conocimiento basada en el Análisis Estadístico Implicativo.

Palabras clave: Análisis Estadístico Implicativo, gestor de conocimiento, investigación universitaria, trabajos de titulación, Escuela Superior Politécnica de Chimborazo, ESPOCH.

Abstract: The Statistical Implicative Analysis allows knowledge management by discovering quasi-implication rules between variables. This research uses the implicative statistics to determine if the university research is in accordance with the research lines and research programs already defined in the Escuela Superior Politécnica de Chimborazo. We are based on the analysis of 970 titling works carried out from 2012 to 2016 in the Sciences Faculty of ESPOCH. We conclude that there is a positive relationship of similarity, cohesion, and implication between titling works, research lines, research programs and the careers. This is a research work replicable and applicable to other educational contexts. This research provides a new knowledge management methodology based on Statistical Implicative Analysis.

Keywords: Statistical Implicative Analysis, knowledge manager, university research, titling works, Escuela Superior Politécnica de Chimborazo, ESPOCH.

Introducción

La Escuela Superior Politécnica de Chimborazo (ESPOCH), “tiene su origen en el Instituto Tecnológico Superior de Chimborazo, creado mediante Ley No.69,09, expedida por el Congreso Nacional, el 18 de abril de 1969, inicia sus actividades académicas el 2 de mayo de 1972” (epoch, 2018). El Instituto de investigaciones IDI es el órgano institucional (IDI, 2018) que se encarga de fomentar y sistematizar la investigación , desde el año 2014 tiene definidas 8 líneas y 33 programas de investigación en vigencia hasta mayo del 2018 (IDI, 2014a). Luego de 3 años de su aplicación se desea conocer si la investigación en la ESPOCH y en particular la investigación de pregrado concuerda con las líneas y programas ya definidos por el IDI (IDI, 2014b). “La Facultad de Ciencias nace en 1978 como Facultad de Química, el 15 de agosto de 1984 se crean las Escuelas de Doctorado en Física y Doctorado en Matemática que junto a la Escuela de Doctorado en Química y Tecnología en Química entran a constituir la Facultad de Ciencias, la carrera de Bioquímica y Farmacia se crea según Resolución No. 311 del Honorable Consejo Politécnico del 7 de Septiembre de 1999 actualmente está conformada por las carreras de Ingeniería Química, Ingeniería ambiental, Química, Física, Biofísica, Matemática, Estadística, Estadística Informática y Bioquímica y Farmacia” (epoch, 2018). Esta investigación se basó en la gestión del conocimiento de todos los trabajos de titulación realizados desde el año 2012 hasta el año 2017 en la Facultad de Ciencias (“DSpace User Registry | DuraSpace,”). La investigación que se realiza actualmente en las universidades ecuatorianas busca transformarse en uno de los principales puntales de la innovación de la economía que tiene actualmente el Ecuador (LOES, 2010; SENPLADES, 2013). El reto constituye entonces contribuir a mejorar las condiciones de vida de la población mediante la investigación (Manosalvas, 2014). Este trabajo pretende relacionar los trabajos de titulación con las líneas de investigación definidos en la Escuela Superior Politécnica de Chimborazo, basados en la clasificación de las áreas del conocimiento de la UNESCO (UNESCO), los Objetivos del Milenio (Ecuador, 2013) y el Plan del Buen Vivir (SENPLADES, 2013) (Caria, Silva, Gallegos, & Domínguez, 2016), dando como resultado una correspondencia usando las técnicas del Análisis Estadístico Implicativo o ASI (del francés *Analyse Statistique Implicative*). El ASI conocido también como Estadística Implicativa (Ritschard, 2005) es una teoría estadística desarrollada en Francia por Regis Gras en los años 80, utiliza los conceptos de similitud, cuasi implicación y cohesión para generar relaciones asimétricas basadas en R-Reglas (Couturier & Pazmiño, 2016). Sus inicios están en la modelización estadística de la cuasi-implicación: cuando la variable o la conjunción de variables “a” es observada en la población, entonces generalmente la variable “b” lo es también. A diferencia de los métodos de análisis simétricos basados en una distancia o en una correlación, los conjuntos de reglas obtenidos pueden conducir a hipótesis de causalidad (Gras, Régnier, & Guillet, 2009). El sistema informático CHIC es un software desarrollado por Raphaël Couturier que automatiza la teoría del ASI (Couturier, 2008). El ASI ha sido utilizado ampliamente en el ámbito educativo (actualmente también en las Analíticas de Aprendizaje (Pazmiño-Maji, García-Peñalvo, & Conde-González, 2016)) y tiene además aplicaciones en las áreas de la psicología, matemática, arte, medicina y otras (Pazmiño, 2014). Desde sus orígenes el ASI ha sido utilizado en la generación y gestión de conocimiento llegando a formalizar su relación con el descubrimiento de conocimiento en bases de datos o KDD (del inglés Knowledge

Discovery in Databases) en el artículo (Pazmiño-Maji, García-Peñalvo, & Conde-González, 2017).

Métodos

La metodología utilizada se ha tomado de (Pazmiño-Maji et al., 2017) y se basa en el proceso del ASI desde el punto de vista del descubrimiento de conocimiento en bases de datos (KDD). La Figura 1 muestra los pasos utilizados.

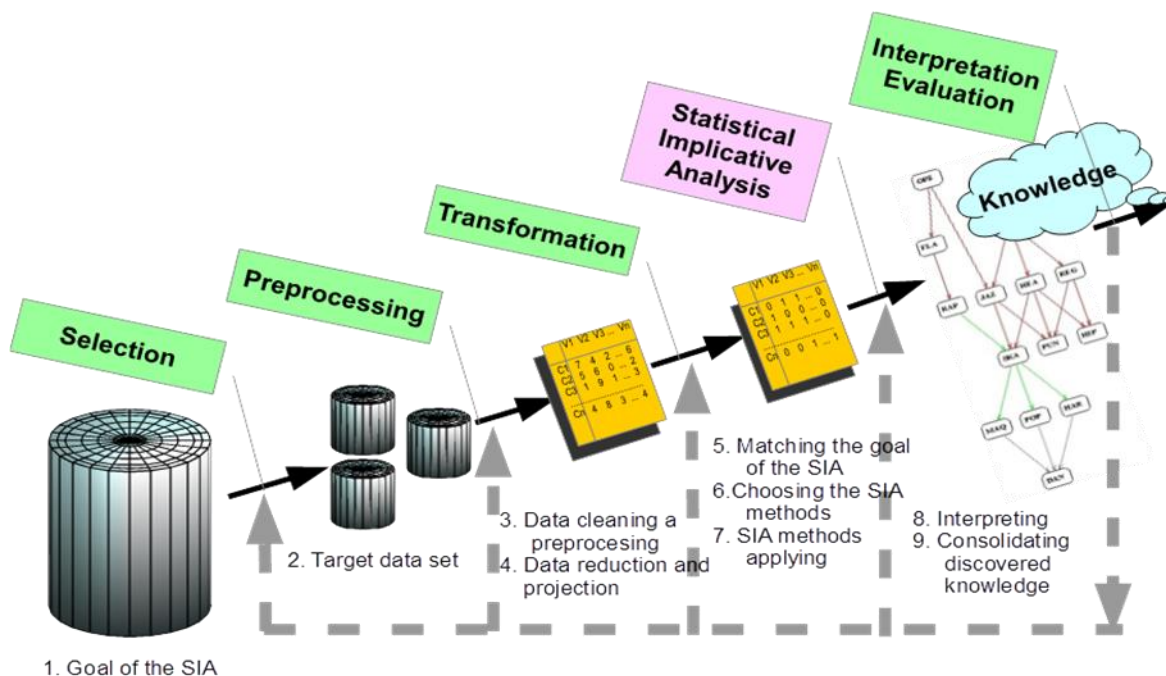


Figura 1: Metodología ASI (Pazmiño-Maji et al., 2017)

A continuación, se detallan los pasos aplicados en la metodología:

Selección: Se utilizaron 970 tesis de la Facultad de Ciencias de la ESPOCH mediante el repositorio abierto de recursos digitales DSpace ("DSpace User Registry | DuraSpace,"). Además, para determinar si la investigación universitaria y en particular la investigación de pregrado de la Facultad de Ciencias está acorde a las líneas y programas ya definidos por el IDI y la ESPOCH. Previamente se investigaron las líneas y programas definidos institucionalmente y luego se elaboraron 35 preguntas guía que ayudaron a asociarlas a los 970 trabajos de titulación analizados. En la Tabla 1, la primera columna muestra en letras mayúsculas las líneas de investigación institucionales y numerados en minúscula los programas de investigación respectivos. En la segunda columna se muestran las preguntas guías asociadas a cada una de las líneas y programas de investigación institucionales. Las preguntas guías fueron elaboradas por un grupo de investigadores del Grupo de investigación Ciencia de Datos de la ESPOCH (CIDED, 2017), especializados en educación y en cada una de las áreas de conocimiento. En la tabla 1, se muestra lo antes explicado.

LÍNEAS	PROGRAMAS	PREGUNTAS
GESTIÓN SUSTENTABLE DE LOS RECURSOS NATURALES Y GESTION DE RIESGOS	1. Biodiversidad sustentable.	¿Según el título, objetivos o hipótesis se propone conservar la diversidad biológica?
		¿Se aplican instrumentos de gestión para salvaguardar los recursos naturales?
		¿La propuesta busca hacer uso de los recursos naturales para satisfacer necesidades con responsabilidad ambiental?
ENERGÍAS ALTERNATIVAS RENOVABLES Y PROTECCIÓN AMBIENTAL.	1. Energías alternativas.	¿Según el título, objetivos o hipótesis se proponen fuentes de energía como alternativa a las tradicionales?
		¿Se aplican técnicas biológicas y/o de ingeniería?
		¿La propuesta contribuye a limitar la emisión de gases de efecto invernadero?
	2. Gestión y tratamiento del aire, agua y suelo.	¿Según el título, objetivos o hipótesis se pretende dar un manejo integral del sistema ambiental?
		¿Se aplican instrumentos de gestión ambiental para la prevención, control y mitigación de impactos ambientales?
		¿Se entrega una herramienta aplicativa?
	3. Biorremediación del ambiente.	¿Según el título, objetivos o hipótesis se pretende mejorar la calidad de los factores ambientales?
		¿Se aplican técnicas biológicas y de ingeniería?
		¿Intervienen organismos vivos o moléculas sintetizadas por ellos para la remediación de agua, aire y/o suelo?
TECNOLOGÍAS DE LA INFORMACIÓN, COMUNICACIÓN, PROCESOS INDUSTRIALES y BIOTECNOLÓGICOS	3. Desarrollo de aplicaciones de software para procesos de gestión y administración pública y privada. Educación;	¿Existen contenidos, metodologías, estrategias educativas o nuevas propuestas pedagógicas que se estudien?
		¿En el título, objetivos o hipótesis se pretende la mejora de la educación?
		¿Se utiliza software para estudiar, mejorar o modificar la enseñanza o el aprendizaje?
	10. Análisis Estadístico Implicativo y computacional	¿Se buscan reglas entre variables?
		¿Existen gráficos de implicación, árboles de cohesión o similaridad?
		¿Se desarrolla software para el ASI o se utiliza el chic o Rchic?
SEGURIDAD Y SOBERANÍA ALIMENTARIA.	1. Evaluación del estado de la seguridad alimentaria.	¿Se evalúa el acceso o la disponibilidad de alimentos?
		¿Se evalúa la inocuidad (que no tenga peligros químicos, físicos o biológicos) y calidad de los alimentos ya existentes?
		¿Se evalúan las condiciones higiénico-sanitarias del lugar donde se elaboran alimentos, aplicando Buenas Prácticas de Manufactura o Prácticas Correctas de Higiene?
	2. Consumo humano para mejorar las condiciones de nutrición y salud.	¿Se investigan la composición química de materias primas no utilizadas convencionalmente?
		¿Se desarrollan productos alimenticios innovadores que mejoren el valor nutricional?
		¿Se evalúa la vida útil de los productos innovadores?
SALUD Y NUTRICIÓN.	1. Evaluación del estado de salud y nutrición.	¿Se mide talla, peso, índice de masa corporal?
		¿Se recolecta información sobre qué alimentos forman parte de la dieta?
		¿Se analizan muestras de sangre, orina, heces?
	6. Administración de servicios de salud, nutrición y alimentación	¿Se evalúa la GESTIÓN de servicios de atención de salud como hospitales, subcentral de salud, farmacias, restaurantes, bares escolares, servicios de alimentación públicos?
		¿Se elabora manuales de calidad para hospitales, subcentros de salud, farmacias, restaurantes, bares escolares, servicios de alimentación públicos?
		¿Se aplican técnicas de la física o biofísica?
	7. Biofísica aplicada a la medicina.	¿Las aplicaciones se relacionan a la medicina?
		¿Existe algún equipo médico como parte del estudio?
		¿Las muestras analizadas corresponden a material vegetal (plantas)?
	9. Desarrollo de fitofármacos.	¿Investigan los componentes del material vegetal?
¿Realizan pruebas de laboratorio del efecto del extracto o de un componente extraído de material vegetal sobre microorganismos, ratones o ratas?		

Tabla 1. Líneas, programas, y preguntas guía para la Facultad de Ciencias (IDI, 2014a)

Preprocesamiento: A las 970 tesis de la Facultad de Ciencias de la ESPOCH, se asoció la información que corresponde a Autor, año, título, carrera, universidad, grado, número de páginas, tutor, fecha, palabras clave, resumen, link, entre otras; para ello se utilizó el gestor bibliográfico EndNote versión 8.0. La ampliación de la información sobre los trabajos de titulación se realizó con un grupo de 5 especialistas de cada una de las carreras que conforman la facultad de Ciencias y que tenían sus tesis en el período de tiempo de estudio (Bioquímica y Farmacia, Biofísica, Física, Matemática, Química, Biotecnología Ambiental, Educación Ambiental, Estadística Informática, Estadística e Ingeniería Química). Se analizaron e ingresaron manualmente los valores cualitativos basándose en las preguntas guías, para ello se utilizó el libro electrónico Microsoft Excel.

Transformación: La base de datos obtenida en el paso de preprocesamiento se transformó en una base de datos dicotómica (Gras & Régnier, 2013) en formato de texto separado por comas (*.csv) conformado por un total de 970 filas correspondientes a los trabajos de titulación y 132 columnas, correspondientes a la información codificada.

Análisis Estadístico Implicativo: Buscando coincidencia entre el objetivo de la investigación y los métodos ASI, se consideraron árboles de similaridad, grafos implicativos y árboles de cohesión. Se leyeron los archivos en formato plano separado por comas y se obtuvieron valores numéricos y gráficos que luego se exportaron en formato JPEG. La aplicación de los métodos ASI se realizó utilizando el software CHIC versión 6.0 (COUTURIER, 2000).

Interpretación y evaluación: La interpretación y evaluación se basan en los grafos implicativos que muestran mediante grafos dirigidos que variables cuasi-implican a cuáles, los árboles de similaridad que muestra las variables más similares utilizando el criterio de similaridad de Israel Lerman (ZAMORA & GREGORI, 2015), el árbol de cohesión que permite ver mediante un dendograma las diferentes relaciones asimétricas entre variables y clase de variables.

Resultados y Discusión

Las siguientes salidas gráficas, muestran ejemplos de la utilización de tres de las herramientas de análisis proporcionadas por el ASI y automatizadas con el software CHIC (Couturier, 2007). Para determinar si la investigación universitaria (en particular la investigación de pregrado) está acorde a las líneas y programas ya definidos por el IDI y la ESPOCH, para lo cual se realizó un Gráfico Implicativo (Régnier, Slimani, Gras, Ben Tarbout, & Dhouibi, 2015) entre todas las carreras y todos los programas de alta utilización en la Facultad de Ciencias. La Tabla 2, muestra la codificación de las variables y los programas de investigación correspondientes para facilitar la interpretación de los gráficos ASI.

VARIABLE	PROGRAMA
T1biodSu	1. Programa de biodiversidad sustentable.
T1enerAlter	1. Programa de energías alternativas.
T2AiAgSu	2. Programa de gestión y tratamiento del aire, agua y suelo.
T3BioAmb	3. Programa de biorremediación del ambiente.

T3SofEdu	3. Programa para el desarrollo de aplicaciones de software para procesos de gestión y administración pública y privada. Educación
T10SIA	10. Programa de análisis estadístico implicative y computacional
T1SegAlim	1. Programa de evaluación del estado de la seguridad alimentaria.
T2ConsHum	2. Consumo humano para mejorar las condiciones de nutrición y salud.
T1EstSalAlim	1. Programa y evaluación del estado de salud y nutrición.
T6AdmSalNutAlim	6. Programa de administración de servicios de salud, nutrición y alimentación
T7BiofMedic	7. Programa de biofísica aplicada a la medicina.
T9Fitofar	9. Programa de desarrollo de fitofármacos.

Tabla 2. Codificación entre variables y programas de investigación respectivamente

En la Figura 2, se observa en la parte derecha que, si un trabajo de titulación está en el programa de energías alternativas, lo compartirá con el programa de gestión y tratamiento del aire, agua y suelo y pertenecerá también a la carrera de Biotecnología Ambiental. En forma similar si un trabajo de graduación pertenece al programa de evaluación del estado de la seguridad alimentaria, también compartirá con el programa de consumo humano para mejorar las condiciones de nutrición y salud, estando en la carrera de Bioquímica y Farmacia. Además, los programas de desarrollo de fitofármacos, consumo humano para mejorar las condiciones de nutrición y salud, administración de servicios de salud, nutrición y alimentación y el programa de evaluación del estado de la seguridad alimentaria están en la carrera de bioquímica y farmacia, desde el punto de vista de su alta utilización.

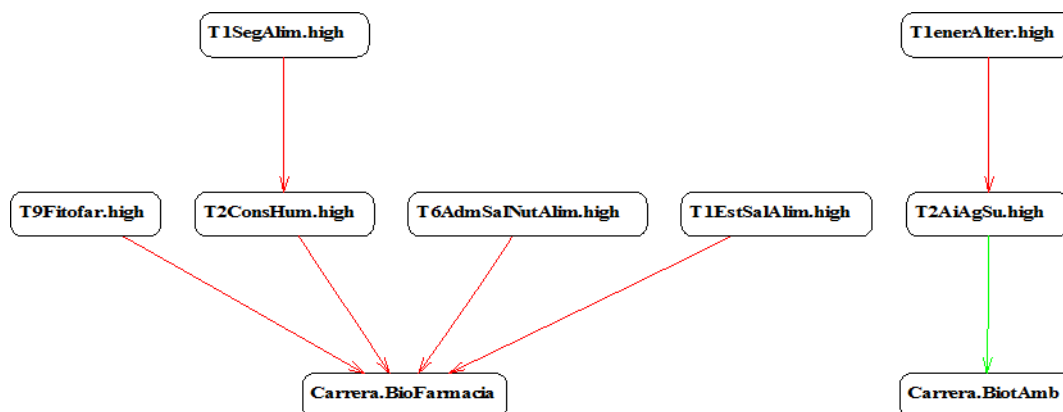


Figura 2: Gráfico Implicative entre los programas de alta utilización en los trabajos de titulación y las diferentes carreras.

El árbol de similaridad (ZAMORA & GREGORI, 2015) de la Figura 3, muestra que el programa de energías alternativas es similar al programa de gestión y tratamiento del aire, agua y suelo en los trabajos de titulación. Además, la línea de color rojo indica que es un nodo significativo, es decir, un nodo correspondiente a una clasificación compatible lo mejor posible con los valores y la calidad del agrupamiento obtenido. Luego en similaridad le siguen el programa uno de seguridad alimentaria (a nivel alto) con el programa dos de consumo humano.

Finalmente, el programa uno de evaluación del estado de salud alimentaria (a nivel alto) tiene una similitud de tercer orden con el programa seis de administración de salud y nutrición alimentaria también a un nivel alto.

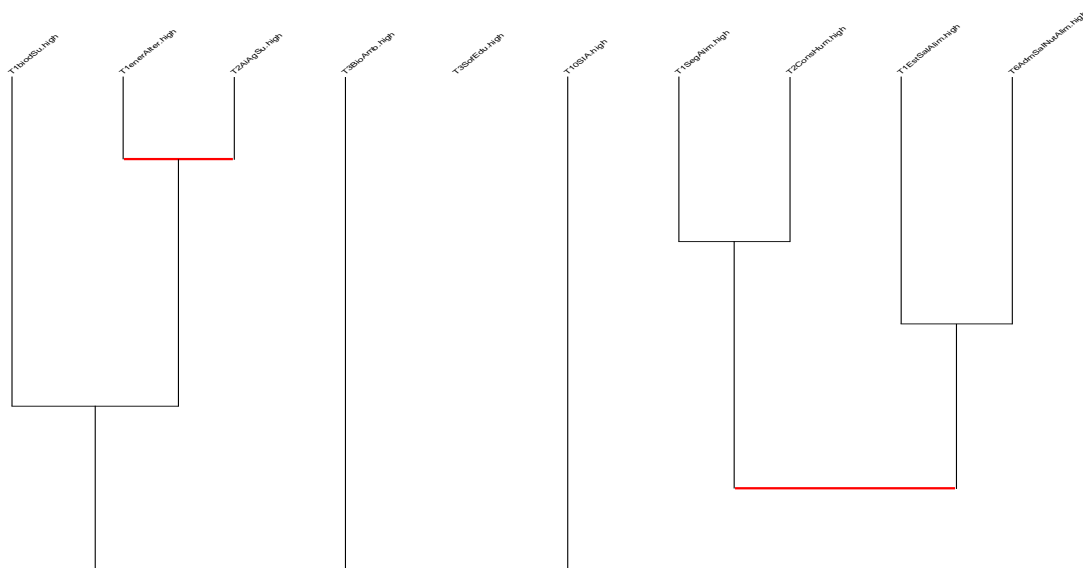


Figura 3: Árbol de Similaridad entre los programas de alta utilización en los trabajos de titulación.

En la Figura 4, se muestra un ejemplo de árbol de cohesión (Regnier & Acioly-Regnier, 2007), el programa de evaluación del estado de la seguridad alimentaria está fuertemente cohesionado (con un valor de 1) con la carrera de Bioquímica y Farmacia desde el punto de vista de su utilización, además son un nodo significativo. De la misma manera, el programa de gestión y tratamiento del aire, agua y suelo está cohesionado (con un valor de 0.961) con la Carrera de Biotecnología ambiental desde el punto de vista de su mediana utilización.

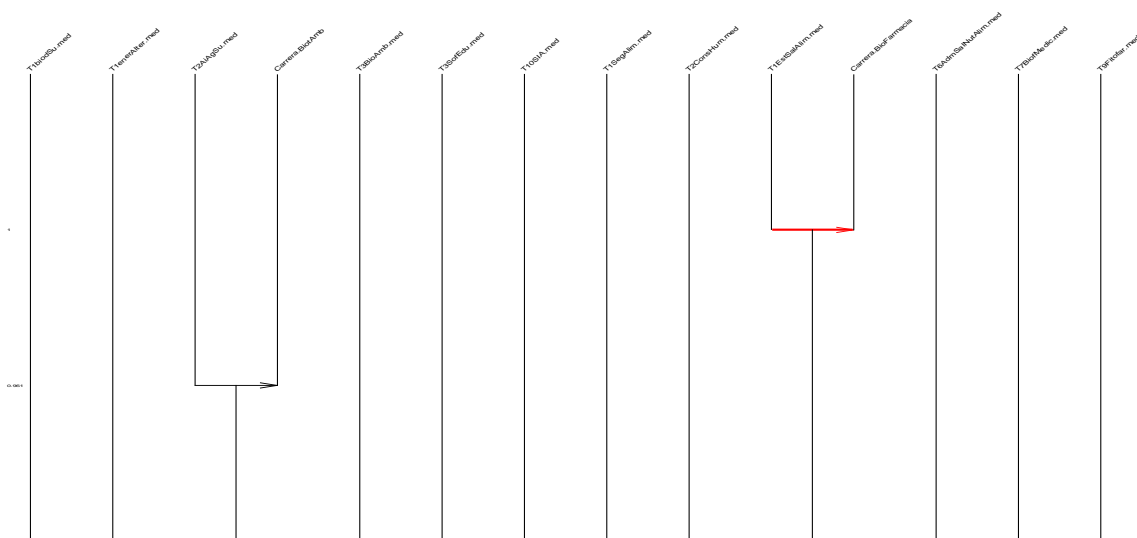


Figura 4: Árbol de Cohesión entre los programas de media utilización en los trabajos de titulación de las carreras de Bioquímica y Farmacia y Gestión ambiental.

Conclusión

El Análisis Estadístico Implicativo permite gestionar fácilmente la información que relaciona las diferentes carreras y su aproximación baja, mediana o alta a los diferentes programas y líneas de investigación propuesta por la ESPOCH. La ejecución de varios gráficos implicativos, similaridad y cohesión nos hace llegar a la afirmación de que hay una relación positiva de concordancia entre los trabajos de titulación y las líneas y programas de investigación definidas institucionalmente. Con una exploración más detallada y exuberante se puede encontrar relaciones de similaridad, cohesión e implicación entre los trabajos de titulación, las diferentes carreras y las líneas y programas de investigación definidas en la Escuela Superior Politécnica de Chimborazo que permitan ver su concordancia con mayor detalle. Se sugiere que se manipulen las diferentes opciones teóricas y de experimentación proporcionadas por el Análisis Estadístico Implicativo y CHIC respectivamente, para llegar a otras conclusiones requeridas por el investigador.

Agradecimientos

A la ESPOCH por el apoyo al proyecto “Análisis e innovación de las investigaciones en la Facultad de Ciencias como estrategia para incrementar la producción científica”. A los investigadores del grupo Ciencia de Datos CIED (CIED, 2017), en especial a Rafaela Pacurucu y Paola Argello por su apoyo en esta investigación.

Referencias

- Caria, Sara, Silva, Verónica, Gallegos, Franklin Ramírez, & Domínguez, Rafael. (2016). Ecuador's Buen vivir. (0094-582X, 1552-678X).
- CIED. (2017, 2017/07/12/03:30:04). Inicio - Grupo de Investigación Ciencia de Datos-CIED, from <http://cided.esPOCH.edu.ec/>
- COUTURIER, Raphaël. (2000). TRAITEMENT DE L'ANALYSE STATISTIQUE DANS CHIC. ASI1, 9.
- Couturier, Raphaël. (2007). CHIC: utilización y funcionalidades. ASI4, 14.
- Couturier, Raphaël. (2008). Chic: Cohesive hierarchical implicative classification Statistical implicative analysis (pp. 41-53): Springer.
- Couturier, Raphaël, & Pazmiño, Rubén. (2016). Use of Statistical Implicative Analysis in Complement of Item Analysis. International Journal of Information and Education Technology, 6(1), 39-43. doi: 10.7763/ijiet.2016.v6.655
- Ecuador, Secretaría Nacional de Planificación y Desarrollo. (2013). Objetivos del Milenio. Retrieved from <http://www.planificacion.gob.ec/wp-content/uploads/downloads/2014/11/Objetivos-del-Milenio-Balance-2013.pdf>.
- esPOCH. (2018, 2017/07/12/03:30:04). Inicio - Escuela Superior Politécnica de Chimborazo, from <https://www.esPOCH.edu.ec/>
- Gras, Régis, & Régnier, Jean-Claude. (2013). Analyse implicative des variables binaires. Intensité implicative. Intensité entropique. In Gras Régis, Régnier Jean-Claude, Marinica Claudia & Guillet Fabrice (Eds.), L'analyse statistique implicative. Méthode exploratoire et

confirmatoire à la recherche de causalités. (pp. 25-46): Cépaduès Editions.

Gras, Régis, Régnier, Jean-Claude, & Guillet, Fabrice. (2009). *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*. Toulouse (France): Cepadues.

IDI. (2014a). PLAN DE INVESTIGACION ESPOCH 2014-2018. In idi (Ed.), (Vol. 2018-10-04). Ecuador.

IDI. (2014b). REGLAMENTO DEL INSTITUTO DE INVESTIGACIONES DE LA ESCUELA SUPERIOR POLITECNICA DE CHIMBORAZO 2014. In El Consejo Politécnico de la Escuela Superior Politécnica de Chimborazo (Ed.). Ecuador.

IDI. (2018). IDI home page, 2018-10-04, from <https://www.esepoch.edu.ec/index.php/investigaci%C3%B3n.html>

LOES. (2010). *Ley Orgánica de Educación Superior*. Quito-Ecuador.

Manosalvas, Margarita. (2014). Buen vivir o sumak kawsay. En busca de nuevos referenciales para la acción pública en Ecuador. Retrieved from <http://revistas.flacsoandes.edu.ec/iconos/article/view/1273>.

Pazmiño-Maji, Rubén A, García-Peñalvo, Francisco J, & Conde-González, Miguel A. (2016). Approximation of statistical implicative analysis to learning analytics: a systematic review. Paper presented at the Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality.

Pazmiño-Maji, Rubén A, García-Peñalvo, Francisco J, & Conde-González, Miguel A. (2017). Statistical Implicative Analysis approximation to KDD and Data Mining: A systematic and mapping review in Knowledge Discovery Database framework.

Pazmiño, R. (2014). Aproximación al Análisis Estadístico Implicativo desde sus Aplicaciones Educativas.

Regnier, Jean-Claude, & Acioly-Regnier, Nadja. (2007). *Analyse cohésitive et interprétations des données dans le champ de l'éducation Nouveaux apports à l'Analyse Statistique Implicative et Applications dans des Disciplines variées* (pp. 329-344): Université JAUME 1 (Castellon) Espagne.

Régnier, Jean-Claude, Slimani, Yahia, Gras, Régis, Ben Tarbout, Ines, & Dhouibi, Ahmed. (2015, 2015). *Analyse statistique implicative. Des sciences dures aux sciences humaines et sociales*. Paper presented at the 8ème Colloque International d'Analyse statistique implicative, Radès Tunisie.

Ritschard, Gilbert. (2005). De l'usage de la statistique implicative dans les arbres de classification. ASI3(11).

SENPLADES. (2013). *Plan Nacional para el Buen Vivir 2013-2017*. (ISBN 978-9942-07-463-8). Quito-Ecuador.

UNESCO. (2018/02/04/03:30:04). Inicio-UNESCO, from <https://es.unesco.org/>

ZAMORA, Larisa, & GREGORI, Pablo. (2015). ALTERNATIVE MEASUREMENTS OF AGREEMENT LEVEL OF INDIVIDUALS TO IMPLICATION AND SIMILARITY FOR CATEGORICAL VARIABLES. ASI8, 18.