



El control de vocabulario como herramienta para mapear intereses de investigación: el caso de las ciencias agrarias en Uruguay

Lucía Simón

Facultad de Información y Comunicación – Universidad de la República, Uruguay

Comisión Sectorial de Investigación Científica – Universidad de la República, Uruguay

lucia.simon@fic.edu.uy

Resumen: Mejorar la calidad de vida de las personas, es fundamental en lo que hace a las cuestiones del desarrollo. Por esto, conocer los intereses de investigación que colaboran con dichas mejoras resulta relevante. Se propone explorar cuáles son estos intereses en ciencias agrarias en Uruguay y en un futuro, estudiar a la solución de qué problemáticas del desarrollo pueden contribuir a resolver. Para alcanzar este objetivo, en este trabajo parcial, se identifican cuáles son los intereses de investigación en ciencias agrarias, mediante técnicas de control de vocabulario.

Existen diversos estudios que estudian cuáles son los temas de investigación en algún ámbito determinado, y en su extensa mayoría, utilizan los CV de quienes investigan como fuente de datos. Mediante el uso de herramientas estadísticas (bibliométricas) se obtienen resultados que pretenden dar cuenta de dichas temáticas. Este trabajo, aporta el uso de una herramienta novedosa para realizar esta tarea: el control de vocabulario (mediante técnicas de indización). Se parte de la premisa, que el control de vocabulario es una herramienta más robusta, con base en dos supuestos: casi ninguna herramienta reconoce sintagmas a la hora de extraer los términos de los textos y, la vinculación de temas de investigación con problemas que éstos contribuyen a resolver es un enfoque novedoso en este tipo de estudios. Finalmente, se presentan las temáticas de investigación de las ciencias agrarias en Uruguay y su posible aporte a la solución de problemas del desarrollo.

Palabras clave: Control de vocabulario; Intereses de investigación; Problemas del desarrollo



1. Introducción

La indagación sobre qué se investiga en un área del conocimiento, país, región, entre otros, se ha realizado a lo largo del tiempo desde diversas perspectivas, en su amplia mayoría a través del uso de currículums vitae (CV) o de publicaciones científicas como fuente de datos y mediante el uso de técnicas estadísticas. Solo por poner un ejemplo, la tesis de Rafael Antunes dos Santos, ya en 2015 mostraba un cuadro con varios estudios sobre la temática (Santos, 2015). Sin embargo, el uso de herramientas propias de la organización del conocimiento (como es la utilización de lenguajes controlados) para precisar este tipo de estudios, ha sido escasamente abordado. De igual modo, la indagación sobre qué problemas pueden ser resueltos a través de las capacidades de investigación existentes (en este caso, a través de los intereses de investigación manifestados), también presenta escasos antecedentes de estudio.

En este sentido, el trabajo presenta, en primer lugar y de forma muy resumida, antecedentes referidos a estudios que han mapeado temas de investigación. A continuación, se presentan algunas consideraciones sobre cuál sería la ventaja de utilizar herramientas como el control de vocabulario para realizar estudios como el que aquí se presenta. Además, se explicitan algunos estudios que trabajaron con las problemáticas a cuya solución pueden contribuir las capacidades nacionales de investigación. Luego se explicitan los pasos que se siguieron para cumplir con los objetivos y se presentan los resultados obtenidos. Finalmente, se esbozan algunas conclusiones referidas a la utilidad y robustez del control de vocabulario (CV) como herramienta para mapear intereses de investigación.

1.1. Antecedentes

A nivel nacional e internacional se encontraron un sinnúmero de trabajos que procuraron realizar análisis similares al propuesto pero con un alcance distinto. El foco, está básicamente en mapear las temáticas en un área determinada, en una revista, en una disciplina, entre otros casos, complementado, muchas veces, por otros análisis. En ese sentido, las principales diferencias, suelen radicar en las fuentes de datos utilizadas y en las técnicas. En particular, los distintos estudios, analizan los temas a través de: i) artículos científicos (aquí es diversa la gama: en algunos casos, es el estudio de resúmenes, títulos y palabras clave, en otros, solo alguna de estas dimensiones, mientras que otro grupo, el análisis integral de los artículos); ii) tesis; producción científica reportada en los CV.

Básicamente, estos trabajos mapean intereses de investigación, mediante el uso de técnicas bibliométricas (análisis de co-palabras o co-ocurrencias y análisis de redes sociales) y, por lo general, el resultado, termina realizando grandes agrupamientos temáticos. En algunos casos, para representar estos agrupamientos, se utilizan categorías establecidas previamente en algún sistema preexistente (grupos de identificación, áreas estratégicas predefinidas, clasificación establecida en el área de interés), o, por la indexación propia de la revista en el caso de trabajos que realizan esta operación, a través del análisis de la producción científica (no de organización del



conocimiento, sino sistemas de clasificación proporcionados por diversos organismos). Dada la cantidad de trabajos en este sentido, se reitera que en Santos, 2015 puede evidenciarse una aproximación a los distintos estudios en este sentido. En este escenario, se intenta evidenciar cuál es el aporte de este trabajo, en relación a los antecedentes encontrados.

2. El control de vocabulario como herramienta para mapear intereses de investigación

El control de vocabulario puede considerarse como una combinación de técnicas y procedimientos que se aplican al lenguaje de forma consistente para establecer expresiones estandarizadas (Barité et al., 2015). En ese sentido, estas expresiones aspiran a ser más adecuadas para la recuperación de información, además, favorecen la comunicación en ámbitos especializados del saber. Esto último, en el entendido de que el lenguaje natural se expresa de variadas formas y puede causar problemas o dificultades a la hora de la comunicación. Puede agregarse que Hjørland (2008) sitúa al control de vocabulario como una de las aproximaciones tradicionales de la Organización del conocimiento, que, como mencionamos parece ser una buena herramienta para desambiguar las inconsistencias del lenguaje natural.

Como se mencionó, el mapeo de intereses de investigación es un tema sobre el que existe un importante número de trabajos realizados. No obstante, las técnicas empleadas para mapearlos, aunque puedan incluir algún tipo de clasificación específica, no necesariamente responden al uso de técnicas de control de vocabulario. Esta afirmación se basa en que, aunque esas clasificaciones puedan ser, en muchos casos, productos de operaciones de control de vocabulario, su uso es marginal, puesto que no se aplican a todos los términos recolectados.

En este marco, los escasos estudios que refieren específicamente al uso de técnicas de control de vocabulario, utilizan esta herramienta en combinación con técnicas bibliométricas (como las mencionadas más arriba). Además, los casos encontrados, responden al área de la Ciencia de la Información (Herrera-Miranda, I.; Licea de Arenas, J. & Gómez-Hernández, J.A. (2013); Montilla-Peña, 2016). Esto último, no parece sorprender, pues el manejo de estas herramientas no es conocido por el resto de la comunidad académica.

Esto último, ha sido una de las motivaciones del trabajo, donde se pretende probar no solo evidenciar la validez de la herramienta para realizar este tipo de estudios, sino que se intenta demostrar la robustez de la misma, en cuánto a la precisión de la información y a su mayor nivel de especificidad. Asimismo, se destaca que el objetivo de estos trabajos, no consistió en probar el uso de la herramienta, sino en hacer otro tipo de análisis, el mapeo de temáticas era uno solo de ellos, para los que se valieron del control de vocabulario con el objetivo de afinar los resultados obtenidos.

Por lo antes mencionado, se entiende que proponer al control de vocabulario como una herramienta teórico-metodológica que permite (de forma precisa) determinar los intereses de investigación en cualquier ámbito especializado del conocimiento, constituye una novedad y un aporte tanto a los estudios vinculados con el comportamiento científico, como a la ciencia de la información.



3. Capacidades de investigación puestas al servicio de la solución de problemas sociales y/o productivos: algunos acercamientos

A efectos de la continuación este trabajo, interesa conocer experiencias que se han centrado en estudiar qué problemas pueden resolverse con las capacidades existentes. Para ello, se destacan: Argenti, Fulgueira y Sutz (1988); Unidad Académica (2006); Boiani (2017) y Cohanoff (2021).

Estos antecedentes buscaron conocer qué problemas del desarrollo pueden resolverse con las capacidades de investigación existentes. Todos tienen en común, que el mecanismo utilizado para recolectar la información fue la encuesta. Sin embargo hay al menos dos diferencias con relación al trabajo propuesto, a saber: i) el universo que aquí se considera es más amplio y se considera más representativo que los restantes, puesto que, para definir quiénes son las personas que investigan en Uruguay, las decisiones tomadas así como las fuentes secundarias utilizadas para su construcción es más abarcativa que la de los estudios precedentes; ii) se consideran “problemas de desarrollo” no sólo los que tienen aplicación en el ámbito empresarial (primer caso), ni los que son solo de aplicación industrial (tercer caso), o que se vinculen a una temática específica, como refería el estudio sobre manejo costero integrado. Si bien, en la encuesta realizada en el año 2006 (segundo caso) la cuestión de los problemas a resolver no estaba orientada a problemas como los antes descritos, el universo aquí considerado podrá, eventualmente, marcar alguna diferencia. No obstante, las tipologías de problemas resultantes de ese estudio serán consideradas/testeadas a la hora de definir las propias.

4. Metodología

Para alcanzar los objetivos propuestos, se diseñó un estudio observacional (Kisch, 2004) de corte cualitativo mediante la aplicación de técnicas de control de vocabulario. La fuente de datos utilizada, fue una encuesta realizada entre agosto y diciembre de 2022 a las personas que investigan en Uruguay. Para este trabajo, se seleccionaron aquellas que indicaron su área de trabajo como ciencias agrarias y se analizaron los campos correspondientes a “intereses de investigación” y “solución de problemas a la que su investigación puede aportar” y se analizó un total de 147 casos, para los cuales se estableció un procedimiento detallado que incluyó los siguientes pasos:

1. Marcado de términos candidatos.

Los mismos, fueron prefijados según dos criterios: su inclusión en el Tesauro AGROVOC; su presunta presencia con fuerte garantía literaria en búsquedas realizadas.

2. Análisis de co-palabras con la fuente de datos establecida: Esta operación se llevo adelante para evitar posibles pérdidas de información.

3. Proceso de normalización: Se realizaron las siguientes operaciones sobre cada término, lo que determinó cambios, exclusiones y alguna incorporación: restricción (en caso de ser necesario y de que la definición existiera en el Tesauro); desambiguación (control de homónimos, términos polisémicos, sinónimos y variantes); jerarquización y relación.



4. Finalmente, se revisó el listado completo obtenido.

5. Primeros resultados y discusión

Se presentan y discuten los resultados obtenidos siguiendo los pasos detallados en la metodología. La población total de la encuesta (es decir, todas las personas que investigan en Uruguay en función de los criterios definidos) suma un total de 3482 personas, de las que 1112 respondieron a esta encuesta (31.9%). A los efectos del análisis del área en cuestión, la población total representaba el 15.5% del total (539 personas), de las que 155 respondieron y cuyas respuestas en los apartados mencionados, fueron objeto de análisis. De las respuestas obtenidas, y para poder analizar cuáles son los intereses de investigación en ese ámbito del saber y a qué tipo de problemas pueden contribuir a su resolución, se chequeó que, al menos, hayan respondido a la pregunta sobre sus intereses de investigación. En este caso, el total de respuestas es de 124 personas. Este es el número de respuestas que fue analizado y sobre el cual se presentan los datos.

La tarea de identificar cuáles son los presuntos intereses de investigación, implicó varias etapas que se desarrollan a continuación. De todos modos, se adelanta que de los diversos análisis, se obtuvo un total de 546 términos que representan los intereses de investigación.

En un primer momento, se confeccionaron tres listados (que se plasmaron en una única plantilla Excel), en función de la procedencia (lenguaje natural de los autores; tesoro AGROVOC; garantía literaria a través de Google scholar) de los presuntos términos a ser representativos de los intereses de investigación. Dichos listados refieren a:

1. Marcado de términos candidatos (manualmente), a través de operaciones de control de vocabulario. Una vez marcados, se contrastaron para fijar su inclusión definitiva con base en dos criterios: su inclusión en el Tesoro AGROVOC; su presunta presencia con fuerte garantía literaria en búsquedas realizadas.

2. Inclusión de las palabras clave proporcionadas por las personas que investigan: Se marcaron los términos que coincidieron con los seleccionados manualmente, se buscaron en AGROVOC (normalización) y en Google scholar.

3. Inclusión de términos derivados de la frecuencia (simple) de las palabras presentes en los campos de análisis de la encuesta: Se marcaron los términos que coincidieron con los marcados inicialmente, se buscaron en AGROVOC (normalización) y en Google scholar.

Aclaración: en los tres casos, se procedió a: i) controlar los términos duplicados; ii) corregir eventuales errores tipográficos; iii) normalizar el uso de plurales y singulares; iv) controlar los sinónimos y v) eliminar las palabras no significativas (este paso, se aplicó para el tercer grupo de palabras seleccionadas, es decir, aquellas que provenían de un listado simple de frecuencias).

Tabla 1. Cantidad de términos seleccionados como posibles intereses de investigación

Tipo de lenguaje	Cantidad de términos	Porcentaje
------------------	----------------------	------------



Control de vocabulario - CV(manual + AGROVOC o garantía literaria)	240	43.9
Palabras clave - PC	77	14.1
Lenguaje natural (frecuencia simple) - LN	37	6.8
CV-PC-LN	18	3.2
CV-PC	37	6.8
CV-LN	129	23.6
PC-LN	8	1.5
Total	546	99.9

Lo primero que se evidencia, es que cerca de la mitad de los términos corresponde solo a un tipo de lenguaje específico: marcado de términos con operaciones de control de vocabulario (el 43.9% del total), mientras que las palabras clave proporcionadas por quienes investigan, representan casi un tercio del total (14.1%). El total de palabras que aparecen, producto del conteo simple, es muy bajo, representando el 6.8% del total.

Por su parte, los términos aportados a través del marcado que se repiten en las proporcionadas por los autores, alcanzan un porcentaje también bajo, pero implican que esos términos, se habrán de fijar para el mapeo de los intereses de investigación. Asimismo, los términos controlados y los términos que aportó la frecuencia simple, son los que alcanzan un porcentaje mayor, pero este resultado no es significativo, ya que, como, es de esperar que los términos que derivan de la frecuencia de palabras, se encuentren en el resto de los lenguajes. No obstante, los términos proporcionados por los autores que coinciden con la frecuencia simple, son 8 (representando el 1.5% del total) y además, son términos generales, que, en todos los casos, tienen un término sintagmático (más específico) asociado.

A continuación, se presentan los resultados obtenidos con relación a la presencia de los términos seleccionados (y normalizados) en el vocabulario controlado AGROVOC y su garantía literaria. Esto último, también se detalla para aquellos términos que no estaban presentes en el lenguaje controlado y se procurará defender su inclusión (o no) en la categorización final, en función de la garantía literaria presente.

En lo que respecta a la garantía literaria, se establecieron dos categorías para determinar su nivel, a saber: 1. Nivel medio de garantía literaria (menos de 500 documento recuperados de



Google scholar, mediante la búsqueda exacta del término); 2. Nivel alto de garantía literaria (más de 500 documentos recuperados de Google scholar con el mismo procedimiento)

Tabla 2. Cantidad de términos presentes en vocabulario controlado y presunta garantía literaria

Tipo de vocabulario	AGROVOC	NO AGROVOC (sólo GL)
Cantidad de términos	442 (81%)	104 (19%)

Tabla 3. Nivel de garantía literaria de los términos presentes en AGROVOC (442)

Nivel de garantía literaria	N.º de términos presentes (AGROVOC)	%
1	33	7.5
2	409	92.5
TOTAL	442	100

Tabla 4. Nivel de garantía literaria de los términos que no se encontraron en AGROVOC

Nivel de garantía literaria	N.º de términos solo con GL	%
1	25	24
2	79	75.9
TOTAL	104	99.9

La tabla 3, evidencia que del total de términos candidatos, una amplísima mayoría tiene respaldo en el vocabulario controlado AGROVOC (81% del total de los términos). Ahora bien, es necesario validar el resto de los términos, y “probar”, que los que se encuentran en el vocabulario controlado, poseen amplio respaldo en la literatura. Sobre este punto, la tabla 4, muestra que el 92.5% de los términos incluidos presentes en AGROVOC, poseen una garantía literaria fuerte.

La tabla 4, muestra que de un total de 104 términos que no están presentes en AGROVOC, un porcentaje menor que los de la tabla anterior (pero igualmente significativo), tiene fuerte presencia de garantía literaria (75.9%). Dicho esto, cabe realizar una aclaración para todos los



casos: cuando el término es demasiado genérico, y claramente escapa exclusivamente al campo de las ciencias agrarias, el resultado de la garantía literaria “se dispara”: esto es razonable, ya que se trata de términos que son abordados desde distintos dominios).

El listado final de los términos, puede descargarse aquí:

<https://docs.google.com/spreadsheets/d/1mOKcf6KGPUffGb4VslaMhgUGC2CezDiJ/edit?usp=sharing&oid=115857080580401682150&rtpof=true&sd=true>

5. Primeras conclusiones

Este trabajo pretendió mostrar todos los pasos seguidos para identificar cuáles son los intereses de investigación en ciencias agrarias en Uruguay. El número de intereses (que podrían ser homologables con temáticas) arrojó un total de 546. Si bien, para este trabajo no se realizó ninguna agrupación jerárquica (que se realizará en el futuro), el nivel de detalle que presentan estos términos, es mucho mayor que los encontrados en la bibliografía que mapea temas de investigación.

El uso de técnicas de control de vocabulario y herramientas derivadas del mismo (en este caso, el vocabulario controlado AGROVOC) demuestra ser robusto para estos fines. Se pretende buscar el modo de agrupar estos intereses en categorías más amplias, pero que puedan evidenciarse cada uno de ellos, pues se considera importante contar con esta información a la hora del diseño de instrumentos relacionados con la política en ciencia y tecnología del país.

Referencias bibliográficas

Argenti, G.; Filgueira, C. y Sutz, J. (1988). *Ciencia y tecnología: un diagnóstico de oportunidad*. Montevideo: MEC-CIESU.

Barité, M., Colombo, S., Blanco, A. D., Simón, L., Cabrera Castromán, G., Odella, M. L. y Vergara, M. (2015). *Diccionario de Organización del Conocimiento: Clasificación, Indización, Terminología* (6ta ed. corr. y aum.). Montevideo: CSIC. Recuperado de: https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/9028/1/barit%C3%A9_diccionario_de_organizaci%C3%B3n_del_conocimiento.pdf

Boiani, M. (2017). *Relevamiento de capacidades de investigación e innovación industrial. Primera Fase. Informe técnico*. Montevideo: CEI.

Cohanoff Liguori, C. (2021). *Capacidades de investigación en manejo costero integrado en Uruguay* (Tesis de maestría). Maldonado: Udelar. CURE. Disponible en: <https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/32804/1/Capacidades%20de%20investigaci%C3%B3n%20en%20manejo%20costero%20integrado%20en%20Uruguay.%20Cohanoff%20Liguori%20Claudia%20Anabel.%202020.pdf>

Herrera-Miranda, I.; Licea de Arenas, J. & Gómez-Hernández, J.A. (2013). *Publicaciones Periódicas en Biblioteconomía, Bibliotecología, Ciencias de la Información y Documentación en México*:



Tendencias Temáticas, Productividad y Redes de Coautoría: 1956-2006. *Revista Interamericana de Bibliotecología*, 36(2), 97-108. Disponible en:
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-09762013000200001&lng=en&tlng=es

Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2-3), 86-101.

Kish, L. (2004). *Statistical Design for Research*. New Jersey: Wiley-Interscience.

Montilla Peña, Leomar José. (2016). Análisis de la producción científica de los artículos de la Revista Zootecnia Tropical del Instituto Nacional de Investigaciones Agrícolas (2006- 2013). *Biblios*, (65), 1-14. Disponible en: <https://dx.doi.org/10.5195/biblios.2016.315>

Santos, Rafael Antunes dos. (2015). *Análise de coocorrência de palavras na pesquisa brasileira em HIV/AIDS indexada na Web Of Science no período 1993-2013* (Tesis de Maestría). Porto Alegre: Universidade Federal do Rio Grande do Sul. Disponible en: <https://lume.ufrgs.br/handle/10183/122180>

Unidad Académica - Comisión Sectorial de Investigación Científica (2006). *Pensando el Plan Estratégico Nacional en Ciencia, Tecnología e Innovación: Elementos para la reflexión derivados de la Encuesta a Docentes en Régimen de Dedicación Total*. Disponible en: https://www.csic.edu.uy/sites/csic/files/documentos/i_pencti_la_universidad_de_la_republica_pien_sa_el_plan_estrategico_nacional.pdf

