

Reclasificación temática de revistas científicas peruanas de Ciencias Sociales y Humanidades

Un análisis léxico-sintáctico y de minería de texto

Por:

Joel Alhuay-Quispe¹; **Josmel Pacheco-Mendoza¹**; **David Quispe²**

1. *Universidad San Ignacio de Loyola, Lima, Perú.*

2. *Universidad Nacional Mayor de San Marcos, Lima, Perú.*



UNIVERSIDAD
SAN IGNACIO
DE LOYOLA



UNIVERSIDAD NACIONAL MAYOR DE
SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA

Agenda

- Situación de general de las revistas científicas peruanas
- Clasificación temática en sistemas de evaluación de revistas
- Clasificación temática de revistas: propuesta de metodología
- Problemas identificados en la metodología propuesta
- Visualización de datos analizados
- Conclusiones
- Referencias

Situación general de revistas científicas peruanas

Revistas por tipo de organización editora

Tipo de organización	N° Revistas	Porcentaje
Universidad	102	73.9%
Asociación profesional	8	5.8%
Instituto de investigación	8	5.8%
Sociedad científica	8	5.8%
Otros (institutos de capacitación, ONG, asociaciones independientes, etc.)	12	8.7%
TOTAL	138	100%

Revistas por área temática general

Área temática general	N° Revistas	Porcentaje
Ciencias	49	35.5%
Multidisciplinarias	42	30.4%
Ciencias sociales	40	29%
Artes y humanidades	7	5.1%
TOTAL	138	100%

Fuente: Santillán-Aldana; Arakaki; de la Vega; Calderón-Carranza & Pacheco-Mendoza. (2017)

Sobre la clasificación de revistas

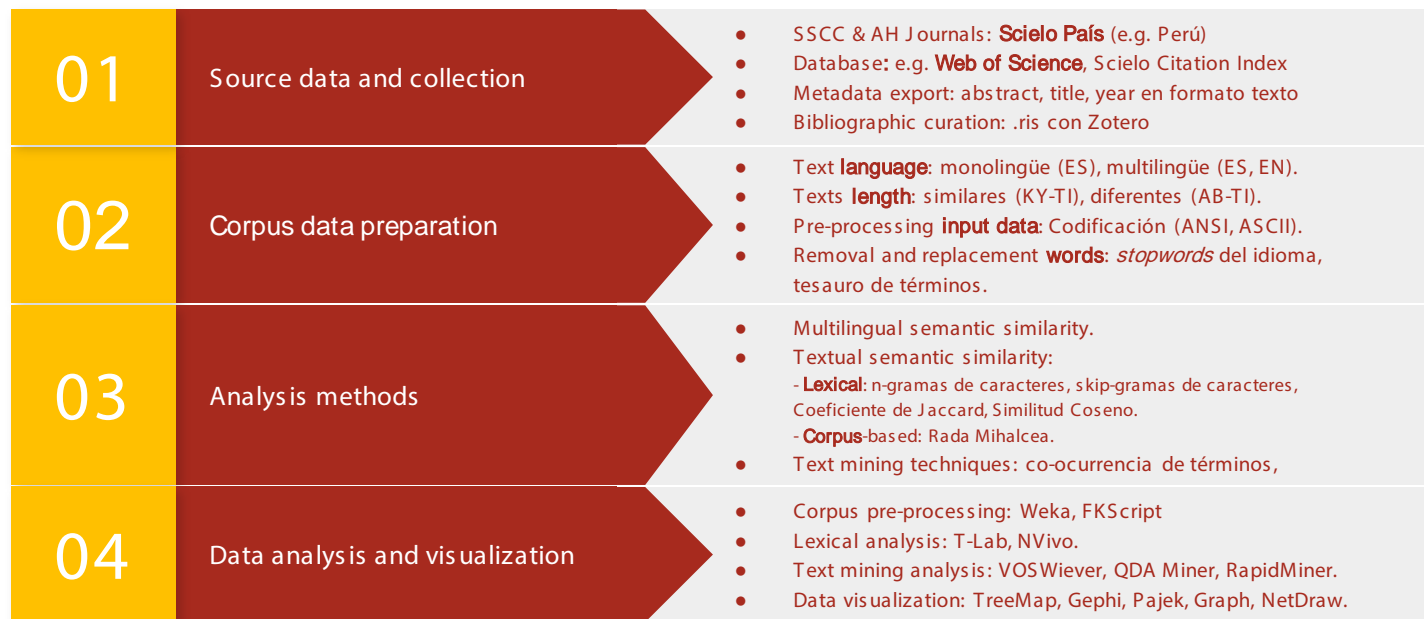
- ❑ El enfoque o clasificación de una revista parte del editor (o editores), pero son las tendencias, los puntos de contacto con otras áreas en común, o similares en su desarrollo, las que van fusionando o ampliando la perspectiva temática de una revista.
- ❑ El posicionamiento y prestigio académico de una revista responde a las áreas del conocimiento en las que fue clasificada; y aunque -generalmente-, existen revistas con pertenencia a una sola área, muchas están catalogadas en más de una.
- ❑ La clasificación o inclusión en determinadas áreas temáticas en las cuales se ubica a una revista, naturalmente influye en el nivel o rigurosidad de evaluación editorial y de calidad de los sistemas e índices de evaluación de revistas científicas.

Sistemas de **clasificación temática** en evaluación de revistas

Fuentes	Áreas	Sub Áreas	Clasificación
Scopus	27	334	<ul style="list-style-type: none">All Science Journal Classification – ASJC - (areas; subject areas)
WoS	22	251	<ul style="list-style-type: none">WoS categories field (CORE) - (subject areas)Essential Science Indicators (ESI) - (areas)
DOAJ	21	181	<ul style="list-style-type: none">Library of Congress Classification (areas; subject areas)
Latindex	7	183	<ul style="list-style-type: none">Library of Congress Classification (areas; subject areas)

Fuente: Sitios web de los sistemas, recogido por los autores.

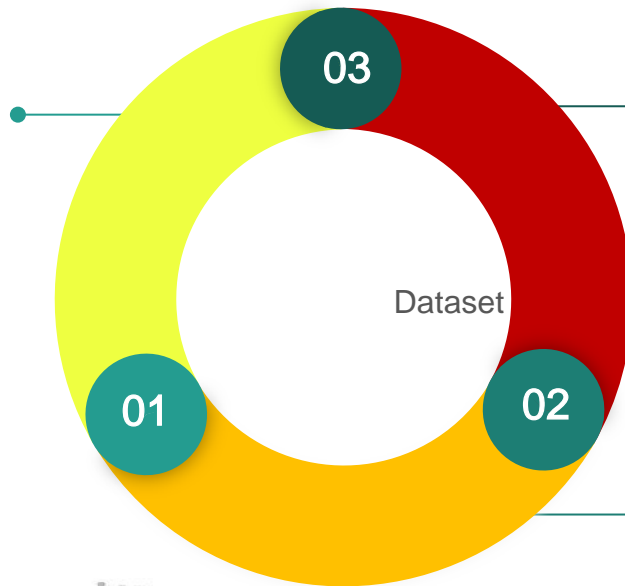
Clasificación temática de revistas: Propuesta de metodología.



Problemas identificados en la metodología propouesta

Collecting data

- **Calidad** de metadata **deficiente** en la fuente (revistas, repositorios).
 - **Predominancia** del idioma (**inglés**) en la base de citas.
 - **Ausencia** de **resúmenes** estructurados.
- **Escaso** uso de **términos controlados** en keywords.



```
<?xml version="1.0" encoding="UTF-8" ?>
<node label="Tesaurus de Biblioteconomia"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="C:\temp\kk\controlledvocabulary.xsd"
  - <isComposedBy>
    <node label="Arxivística" />
    <node label="Biblioteconomia" />
    - <isComposedBy>
      <node label="Accés al document bibliotecari" />
      <node label="Automatització de Biblioteques" />
    </isComposedBy>
  </node>
  <node label="Ciències i tècniques auxiliars" />
  <node label="Estudis mètrics de la informació" />
  <node label="Fonts d'informació" />
  </isComposedBy>
</node>
```

Metadata

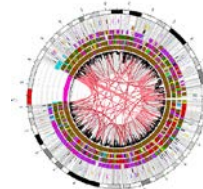


Data analysis

- Normalización compleja de términos respecto a disciplinas (e.g. OCDE)
- Software de análisis predominantes en inglés.
- Entrada y salida de datos en formato no estándar.

Term	Occurrences	Balance
study	34	0.29
research	34	0.13
paper	13	0.11
analysis	16	0.13
effect	27	0.17
result	36	0.17
evidence	17	0.19
finding	17	0.20
author	14	0.21
journal	16	0.22
article	16	0.22
cost	8	0.28
year	6	0.29
increase	17	0.29
impact	16	0.33

Dataset



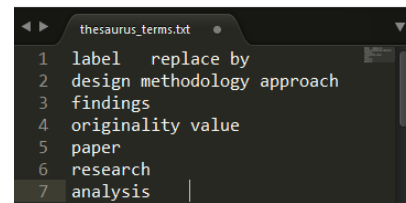
Visualization

Preparing the corpus data

- Datos bibliográficos de diferente longitud.
- Corpus textual en más de un idioma.
- Exclusión de términos vacíos no re-utilizables.



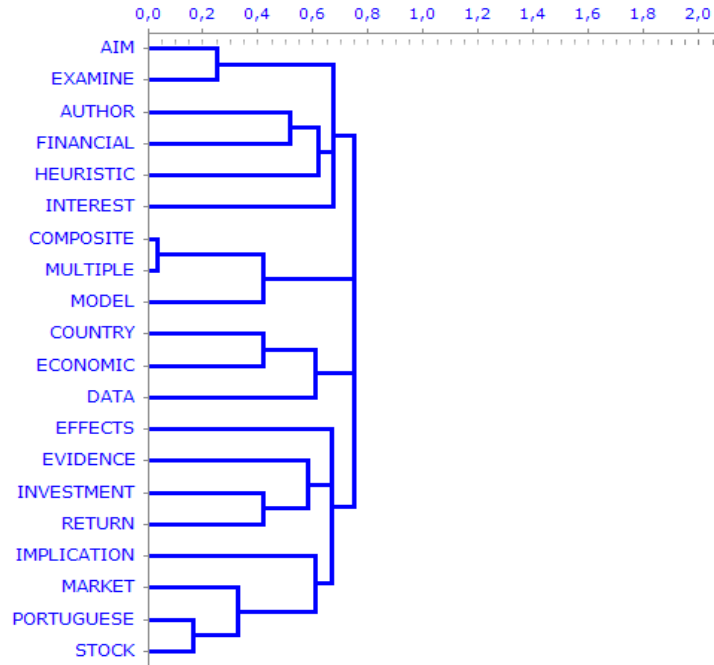
Corpora



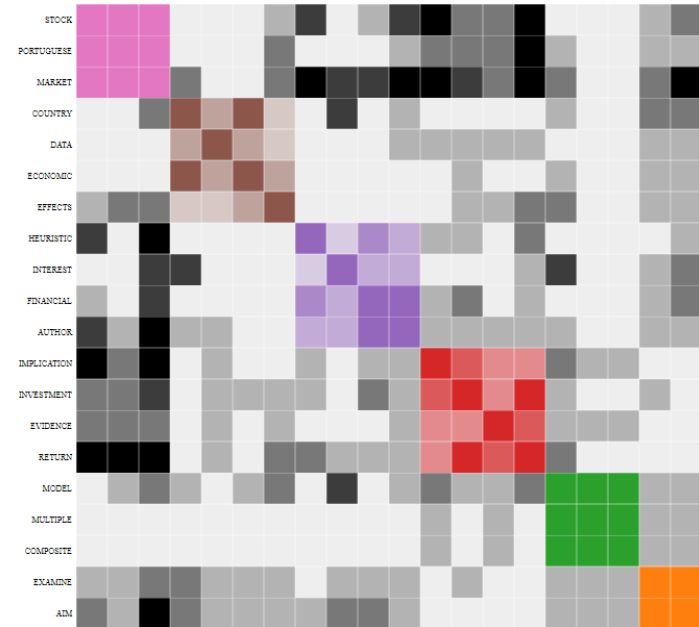
Thesaurus

Visualización de datos analizados: análisis de prueba

Dendrogram clustering



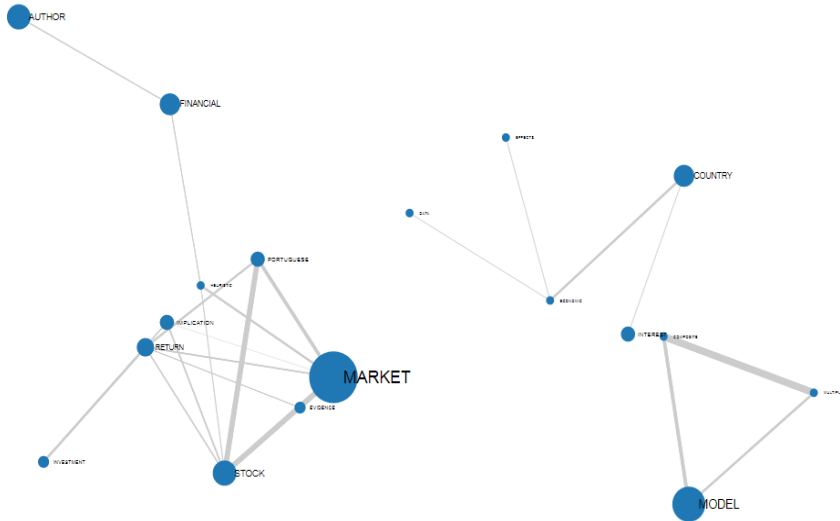
Co-Occurrence Matrix



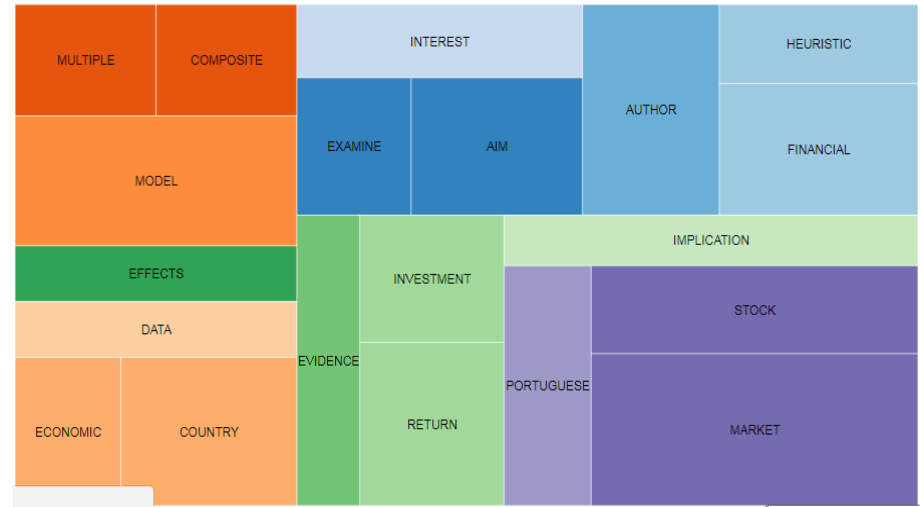
Data: The Journal of Economics, Finance and Administrative Science - JEFAS, muestra año 2017.

Visualización de datos analizados: análisis de prueba

Network Analysis



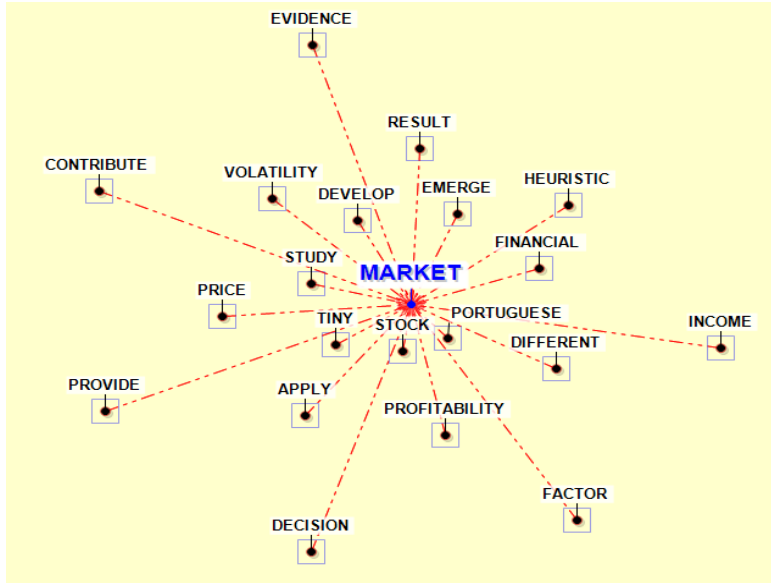
Treemap clustering



Data: The Journal of Economics, Finance and Administrative Science - JEFAS, muestra año 2017.

Visualización de datos analizados: análisis elegidos

Diagrama radial

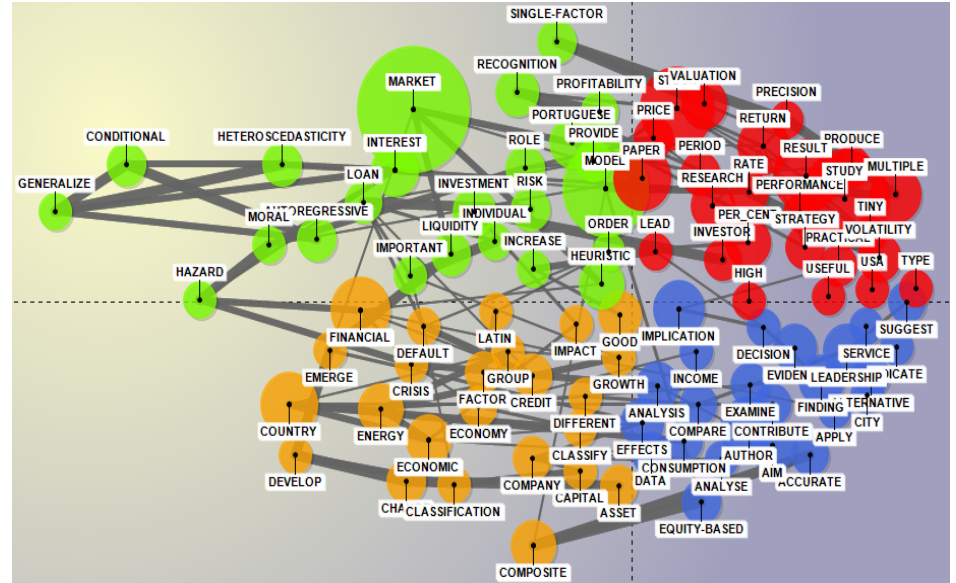


índice: Jaccard.

Contexto: n-gramas=2.

Método: Asociaciones de palabras

Escalamiento multidimensional



índice: Coseno.

Contexto: n-gramas=2.

Método: Sammon.

Conclusiones

- La **minería de textos** sobre datos bibliográficos permiten mostrar el **estado de un área de conocimiento**; sin embargo, también permite establecer lineamientos de **clasificación temática** empleando y combinando técnicas y métodos basados en **datos textuales** de los artículos (**metadatos**).
- El uso del método de **Escalamiento multidimensional** permitirá a los editores conocer, dependiendo de las ventanas de tiempo, cuáles son los tópicos: **Emergentes, Estables, Decrecientes y Marginales**.
- Una **clasificación** más **precisa** de las revistas, en los índices y sistemas de evaluación, repercutirá a favor de la **mejora** en la **visibilidad** o **posicionamiento** de una revista científica en el campo que publica.