

**REPRESENTACIÓN DEL CONOCIMIENTO EN HUMANIDADES: modelado conceptual de obras depositadas en el Repositorio Institucional de la Universidad Federal de Bahía - Brasil**

***REPRESENTATION OF KNOWLEDGE IN HUMANITIES: conceptual modeling of works deposited in the Institutional Repository of the Federal University of Bahia – Brazil***

***REPRESENTAÇÃO DO CONHECIMENTO EM HUMANIDADES: modelagem conceitual das obras depositadas no Repositório Institucional da Universidade Federal da Bahia - Brasil.***

**Bruna Lessa**

Universidade Federal da Bahia, Brasil

[brunalessa@ufba.br](mailto:brunalessa@ufba.br)

<https://orcid.org/0000-0003-4485-203X>

**Tauane Pereira Sales**

Universidade Federal da Bahia, Brasil

[taysales6@gmail.com](mailto:taysales6@gmail.com)

<https://orcid.org/0009-0009-0386-6559>

**Jeane Silva de Carvalho**

Universidade Federal da Bahia, Brasil

[jeane.carvalho@ufba.br](mailto:jeane.carvalho@ufba.br)

<https://orcid.org/0009-0008-3044-8259>

**Resumen**

Este trabajo trata sobre el desarrollo de procedimientos de recolección, extracción y visualización de conceptos como metodología para la creación de vocabulario controlado para las áreas de Humanidades, del Repositorio Institucional de la Universidad Federal de Bahía, Brasil. Tiene un carácter exploratorio y experimental, en cuanto al uso de redes conceptuales, con base en los fundamentos teóricos de la Organización del Conocimiento y la Teoría de Redes, y el enfoque del Análisis de Redes Sociales, para una representación del conocimiento basada en tipos semánticos. Para la recolección y extracción de datos se utilizó Python, y el *software Gephi* para la creación y análisis de redes semánticas. El resultado es la visualización de la interconexión entre conceptos de

las áreas del campo de investigación en Humanidades. Se considera que la relación entre estos conceptos facilite la recuperación de las obras depositadas en el Repositorio.

**Palabras clave:** Representación del Conocimiento, Humanidades Digitales, Repositorio Institucional – Brasil, Análisis relacional, Modelización.

### **Abstract**

This work deals with the development of procedures for collecting, extracting, and visualizing concepts as a methodology for creating controlled vocabulary for the areas of Humanities, from the Institutional Repository, of the Federal University of Bahia, Brazil. It has an exploratory and experimental character, regarding the use of conceptual networks, based on the theoretical foundations of Knowledge Organization and Network Theory, and the approach of Social Network Analysis, for a representation of knowledge based on semantic types. For data collection and extraction, Python was used, and the Gephi software for the creation of the semantic network and analysis. The result is the visualization of the interconnection between concepts of the areas of the field of research in Humanities. It is considered that the relationship between these concepts facilitates the recovery of the works deposited in the Repository.

**Keywords:** Knowledge Representation, Digital Humanities, Institutional Repository – Brazil, Relational Analysis, Modeling.

### **Resumo**

Este trabalho trata do desenvolvimento de procedimentos de coleta, extração e visualização de conceitos como metodologia para a criação de vocabulário controlado para as áreas de Humanidades, a partir do Repositório Institucional da Universidade Federal da Bahia, Brasil. Tem um carácter exploratório e experimental, em termos da utilização de redes conceituais, com base nos fundamentos teóricos da Organização do Conhecimento e da Teoria das Redes, e da abordagem da Análise de Redes Sociais, para uma representação do conhecimento baseada em tipos semânticos. Para a coleta e extração de dados foi utilizada a linguagem *Python* e para a criação e análise das redes semânticas foi utilizado o *software Gephi*. O resultado é a visualização da interconexão entre conceitos das áreas de investigação das Humanidades. Considera-se que a relação entre esses conceitos facilitará a recuperação das obras depositadas no Repositório.

**Palavras-chave:** Representação do Conhecimento, Humanidades Digitais, Repositório Institucional - Brasil, Análise Relacional, Modelage

## Introducción

Las relaciones entre el campo de las Humanidades Digitales y las Ciencias de la Información se pueden identificar de varias maneras, ya que ambos tienen adherencia investigativa en estudios sobre organización, acceso y uso de la información. En particular, en el campo de la Organización del Conocimiento (KO), que involucra teorías y prácticas relacionadas con la modelización del conocimiento para favorecer el desarrollo de representaciones sobre un determinado dominio del conocimiento, como los Sistemas de Organización del Conocimiento (SOC), que permiten modelar la estructura semántica de un dominio de conocimiento para organizar la información y promover la gestión del conocimiento, relacionar las prácticas y herramientas utilizadas en Humanidades Digitales en la aplicación de tecnologías con objetivos de análisis de grandes conjuntos de datos en las áreas de Ciencias Humanas y Ciencias Sociales Aplicadas, contribuyendo a una mejor comprensión de la cultura humana y el conocimiento en el mundo digital actual.

Inusualmente, en el modelado de datos, el uso de redes semánticas sirve para mapear la frecuencia e intensidad de relaciones entre entidades (sean términos y/o instituciones, por ejemplo) y, específicamente en KO, bajo el sesgo de teorías constituidas en el contexto de la Biblioteconomía y Documentación, se ha utilizado como sistemas donde los conceptos se modelan como nodos en una red de tipos de relaciones variables, facilitando a través de herramientas digitales la creación de gráficos y diagramas que muestran las relaciones entre términos de manera visual e interactiva, el análisis e interpretación de datos en diversas áreas del conocimiento.

En este contexto, el estudio de las relaciones entre las palabras y sus conceptos se puede aplicar a cualquier tipo de documento, ya sea físico o virtual. En artículos, por ejemplo, puede identificar las similitudes entre diferentes tipos de métodos y teorías aplicadas en diferentes dominios del conocimiento, lo que permite percibir las características del objeto de investigación y el enfoque conceptual de su aplicación. Así, este artículo presenta el resultado de la primera fase del proyecto de investigación "*Modelado de redes conceptuales como metodología para la creación de vocabulario controlado para la Universidad Federal de Bahía - Brasil*", que tuvo como objetivo desarrollar procedimientos para la recolección, extracción y visualización de conceptos como metodología para la creación de un vocabulario controlado en Humanidades, a partir de obras depositadas en el Repositorio Institucional de la Universidad Federal de Bahía-Brasil.

Actualmente, los sistemas de recuperación de información para elementos documentales producidos en la Universidad Federal de Bahía no tienen un vocabulario controlado que permita una recuperación eficiente para diferentes propósitos. Así, esta investigación exploratoria y experimental entiende la necesidad de ampliar el conocimiento sobre las características y potencialidades de estos instrumentos, sobre todo, de manera interdisciplinaria entre las áreas de

Ciencias de la Información, Biblioteconomía y Humanidades Digitales, ya que el estudio de las redes conceptuales y semánticas es transversal en estas áreas, ya sea de forma teórica, práctica o aplicada.

Como resultado preliminar, se cree que el modelo conceptual permitirá desarrollar un vocabulario institucional controlado, cuyo significado es representar los elementos documentales generados en la Universidad, pertenecientes a su patrimonio bibliográfico, iconográfico y documental, registrados en sistemas y bases de datos, como el Repositorio Institucional, colaborando para la organización, clasificación y recuperación de la información científica.

### **Modelos conceptuales para la representación de la información**

Las investigaciones realizadas a mediados de la década de 1990 y principios de la década de 2000 privilegiaron el estudio sistemático de las teorías subyacentes a las Ciencias de la Información, la Terminología y Ciencias de la Computación (CC), relacionadas con la construcción de modelos conceptuales, que permitieron la elaboración de lenguajes documentales, sistemas computacionales, hipertextos, sistemas destinados a construir bases de conocimiento - llamadas sistemas expertos - y, más recientemente, en el campo de la inteligencia artificial, las denominadas ontologías.

Para Campos (2004), existen puntos de investigación sistemática de los principios fundamentales del modelado de dominios de conocimiento a través de métodos y teorías, con la perspectiva de dotar al profesional de la información de un aparato de herramientas representacionales, enumerando cuatro de estos principios que pueden ser utilizados en el proceso de modelado, especialmente en los campos de las CI, la Terminología y CC.

El primer principio se refiere al método de razonamiento utilizado para organizar el conocimiento dentro de un dominio; la segunda analiza cómo se define el objeto de representación, es decir, cuál es la unidad de conocimiento que será representada; el tercero se refiere a la relación entre los objetos, con el objetivo de verificar las posibilidades de conexión/separación semántica entre los conceptos de un dominio dado; y el cuarto destaca las formas de representación gráfica que puede adoptar un modelo (Campos, 2004).

El proceso de modelado requiere el desplazamiento del “mundo fenoménico” a un espacio de representación. Se entiende que el concepto de representación del conocimiento en el proceso de modelado está relacionado con los roles que puede jugar la representación (Davis *et al.*, 1992). De esta forma, los mecanismos de representación del conocimiento permiten representar fácilmente los procesos de formalización sobre los objetos y sus relaciones, en contextos predefinidos.

Desde una perspectiva estructural, Brachman (1979) explica que la representación del conocimiento se puede clasificar según los tipos de primitivas que se ofrecen al usuario, en cuatro niveles: lógico, epistemológico, ontológico y conceptual. Desde la perspectiva de CI y CC, por ejemplo, los niveles que nos

interesan son aquellos en los que se puede sistematizar y representar una estructura de conocimiento a partir de contextos específicos. Según Newel (1982), los niveles epistemológico y ontológico son los que permiten la representación del conocimiento estructurado y formalizado.

### **Sistemas de organización del conocimiento y su aplicabilidad en el proceso de representación y recuperación de la información**

Si bien se comprende la necesidad de la persona humana de representar para recuperar información, siendo esta necesidad la motivación de las reflexiones filosóficas que fundamentaron la construcción de teorías y prácticas de Organización del Conocimiento y Representación de la Información, se comprende la necesidad de la interdisciplinariedad y la interoperabilidad en el uso de las técnicas de representación de la información y los métodos de recuperación de la información aliados a las nuevas tecnologías, que permitan el desarrollo de modelos conceptuales que permitan la creación de un vocabulario controlado que satisfaga las necesidades de recuperación del conocimiento y de la información en un dominio de conocimiento.

Ante el panorama de la organización del conocimiento y la información y sus variaciones conceptuales y construcciones terminológicas, en el ámbito de la Biblioteconomía y la Información (BCI), los aportes de Ranganathan y su Teoría del Análisis Facetado (TAF), y de Dahlberg, a través de la La Teoría de Conceptos (TC), con sus fundamentos teórico-metodológicos, es la más utilizada en el modelado y representación de información. Otros estudios corroboran la perspectiva de la necesidad de interoperabilidad e interdisciplinariedad entre los diferentes dominios del conocimiento, como las alternativas teórico-metodológicas que surgen de la Lingüística, y que parten de una base semántica del estudio del lenguaje, por ejemplo, la Teoría de la Valencia, la Gramática de Casos, la Gráfica Conceptual y la Teoría de la Gramática Funcional, explorada en detalle por los autores Café y Brascher (2011).

Después de todo, no es posible obtener una organización del conocimiento funcional sin considerar la relevancia de los aspectos semánticos de las relaciones conceptuales involucradas en estos estudios. Un interrogante cobra relevancia al relacionar la referida investigación en proceso de construcción conceptual y fundamentación teórico-metodológica a lo que presenta la propuesta de enfoques orientados al estudio sintáctico-semántico del lenguaje a la construcción de metodologías con el fin de representar dominios de conocimiento, que se trataría de la posibilidad de una visualización y representación de conceptos completos, planteada por Campos (2004), o que extrapolan los límites ontológicos abordados por Lima (2020) para facilitar el proceso de recuperación de información por parte del usuario, principalmente en la *web*. Aunque no respondida por Souza, Tudhope y Almeida (2012) en sus estudios que exploran la aplicabilidad del SOC, la pregunta planteada por los autores complementa la comprensión de que las

dimensiones de clasificación del SOC serían una forma de representar gráficamente conceptos complejos (como la clasificación taxonómica).

Por supuesto, los enfoques metodológicos para el desarrollo de SOC no pueden limitarse únicamente al campo de BCI. Con vistas a la comprensión de la complejidad del modelado para la representación de la información, se cita el *Modelo Dinámico de Análisis Conceptual* (Modac) propuesto por Novo y Miranda (2015), que es una estructura más de organización del conocimiento, formando un SOC dinámico, así como las teorías que sustentan tales conceptos. El modelo presenta características de otras estructuras como la taxonomía y la ontología, sin embargo es más amplio porque no se limita a la base de modelos que analizan solo palabras clave, sino un conjunto de frases (texto) que define un pensamiento presentado por un autor en su investigación, explorando así múltiples posibilidades del ser, aunque el objetivo sea condensarlo en una propiedad general (término) para representarlo mejor, y en el futuro ordenarlo para recuperarlo con mayor precisión.

## **Representación de la información y construcción de redes semánticas**

Redes semánticas son sistemas donde los conceptos se modelan como nodos en una red de diferentes tipos de relaciones. En la representación de la información, las redes semánticas son bidimensionales, es decir, la relación entre los objetos y/o conceptos analizados puede tener más de un tipo de relación, estructura básica para organizar el conocimiento en redes. Son más ricos que los tesauros en la definición de categorías o relaciones semánticas y “[...] representa el conocimiento en forma de nodos (conceptos, atributos) conectados por arcos (relaciones)” (Sales, Campos y Gomes, 2008, p.64).

Aparecen como una adaptación del tesoro conceptual, con el objetivo de obtener una mejor visualización de la relación entre conceptos, posibilitando la navegación a través de grafos de conexión y así representar el conocimiento. En general, estas relaciones semántica están representados por grafos integrados por los conceptos (nodos) y por los arcos (relaciones) que interconectan los conceptos, y esta conexión puede representar propiedades entre estos conceptos, configurándose así como un avance del SOC en relación a la proximidad de los términos del lenguaje natural. Para Novak y Cañas (2009), en estas redes los nodos representan los conceptos, mientras que los arcos, las relaciones que se dan entre los conceptos, con un impacto visual de mayor precisión en relación a los conceptos. los de jerarquía.

El concepto de redes semánticas fue utilizado por primera vez en 1967 por Ross Quillian , sentando las bases de su proyecto que contemplaba la estructuración de la memoria semántica del ser humano con un esquema formal a través de un mecanismo asociativo en el que era posible encontrar palabras, tal como lo presenta Brachman (1979). Antes de eso, en la década de 1950 , se crearon inicialmente las redes semánticas , para que las computadoras pudieran traducir Lenguaje natural automático.

El esquema de Quillian , la red estaba representada por palabras (nodos) y los vínculos entre ellas (bordes), a partir de información asociada, lo que contribuyó a que los estudios centrados en las redes semánticas, en la década de 1970, tuvieran como énfasis un enfoque centrado en la estructura del conocimiento.

Sin embargo, fue a partir de los estudios de Woods (1975) que se puede encontrar la base teórica para comprender la estructura de las redes semánticas . La noción de *vínculo* , o punto de conexión , por ejemplo, característica estructural de una red, permite comprender cómo se representan los vínculos entre conceptos , expresando así las relaciones semántica entre los términos analizados. El estudio de las relaciones entre las palabras y sus conceptos se puede aplicar a cualquier tipo de documento, ya sea físico o virtual. En artículos, por ejemplo, puede identificar las similitudes entre diferentes tipos de métodos y teorías aplicadas en diferentes dominios del conocimiento, lo que permite percibir características del objeto de investigación y el enfoque conceptual de su aplicación.

En el contexto de la Inteligencia Artificial, las redes semánticas se utilizan en casi todas las áreas de aplicación, incluida la comprensión del lenguaje natural, bases de datos deductivas, recuperación de documentos en bibliotecas digitales, planificación empresarial, diagnósticos médico, análisis de casos judiciales, razonamiento analógico, sistemas expertos, robótica, entre otros.

En el campo de las Ciencias de la Computación , la construcción de redes semánticas se basa en la Teoría de Redes, y sus principales autores se basan en teórico Freeman (1978), Wasserman y Faust (1994), Watts y Strogatz (1998), Watts (1999), Strogatz (2001), Albert y Barabasi (2002), Newman (2003), entre otros. Estos autores abordan definiciones , métricas y metodologías que involucran el análisis de redes sociales, con base en la Teoría de Grafos .

En Ciencias de la Información, se ha utilizado en el campo de la Organización del Conocimiento como un SOC para organizar conjuntos de términos que representan conceptos, modelados como los nodos de una red con tipos de relaciones variables. Los tipos de relaciones más comunes son:

- a) “ es-un ” / “es un” : relación jerárquica entre clases ( hombre - es un - ser humano);
- b) “ es un instancia de ” / “es una instancia de” : relación entre instancia y clase (el gato de John – es una instancia de – gato);
- c) “ es parte de ” / “es parte de” : relación entre parte y todo (fin – es parte de –pez);
- d) “ tiene ” / “tem” : relación entre objeto y atributo (gato – tiene – cola).

En el contexto de la modelación conceptual, entendida en este estudio como una de las técnicas utilizadas en el campo de la CO, en el proceso de representación de las relaciones entre conceptos y/o entidades, en un determinado dominio de conocimiento, uno de sus objetivos es favorecer el análisis de una información, un problema o incluso un proceso, de forma más clara y organizada, además de la visualización a partir de gráficos, por ejemplo.

Al relacionar, por lo tanto, la técnica de modelado conceptual y el enfoque de análisis de redes sociales (ARS), es posible mapear conceptos y flujos de información y conocimiento en un dominio de conocimiento. Tal modelado, producido bajo relaciones semánticas, representa las relaciones conceptuales entre información textual, por ejemplo, de títulos, palabras clave, lenguaje natural utilizado en un texto científico. Esta técnica permite visualizar las relaciones de jerarquía, asociación, parte-todo, entre otras, permitiendo otros tipos de modelado, que pueden permitir, por ejemplo, describir relaciones entre los diversos conceptos de un área específica del conocimiento, como las ontologías.

En la siguiente sección, se presenta la metodología utilizada en esta investigación, seguida de la presentación de datos y resultados.

## Metodología

En el proceso metodológico de esta investigación, tratamos de encontrar en los estudios sobre SRI una metodología que fuera más adecuada para representar los dominios del conocimiento más allá del dominio de las Ciencias de la Información, buscando una aplicación que contemplara otras áreas como las Ciencias de la Computación, Ciencias de la Información, y Ciencias Lingüísticas, culminando así en un enfoque metodológico interdisciplinario, que pretende sumar esfuerzos teóricos y metodológicos para validar el proceso de modelación.

El universo de este estudio es la producción científica organizada en el Repositorio Institucional de la Universidad Federal de Bahía (UFBA), Brasil, tomando como muestra la producción científica en el campo de las Humanidades.

El desarrollo de esta investigación, de carácter exploratorio y descriptivo, puede considerarse también como experimental. Su naturaleza es aplicada, ya que a partir de un fenómeno real se pretende presentar el enfoque del análisis de redes sociales para la visualización de conceptos como procedimiento metodológico para la creación de vocabulario controlado.

Los procedimientos técnicos de recolección y análisis de datos se basaron en dos implementaciones: la primera conceptual y la segunda tecnológica. Para el conceptual, de las áreas de conocimiento en Humanidades, se utilizó vocabulario controlado específico para su análisis temático y clasificación. Para la tecnología se utilizó *el software Gephi y la librería Python*, así como bases teóricas de las Ciencias de la Computación, enfocadas al análisis de redes sociales complejas.

Así, el proceso se dividirá en tres fases:

- 1 – Extracción de datos del Repositorio UFBA
- 2 - Creación de una base de datos textual
- 3 - Minería de textos de trabajos académicos para crear la red conceptual

Para tales fases, fue necesario tener una comprensión teórica de la Teoría de la Clasificación Facetada (Ranganathan), la Teoría de los Conceptos (Dahlberg, 1978), la Teoría de Redes (Wasserman y Faust; Albert y Barabasi; Newman), y el enfoque del Análisis de Redes Sociales (Watts y Strogatz). *El software utilizado cumplió con el proceso de agrupamiento*, que involucra identificación y similitud

de conceptos, es decir, un programa escrito para crear redes de palabras, y otro para visualización de redes, los cuales fueron elegidos a partir del análisis del corpus y contexto de los datos extraídos, con el objetivo de su mejor adecuación a la muestra de investigación.

### **Extracción y análisis de datos en el repositorio UFBA**

Para el tratamiento de los datos del área de Humanidades en el Repositorio de la UFBA, se establecieron criterios para la extracción de los metadatos vinculados a los ítems analizados: tipo (tipología del documento – artículo/tesis/disertación), título, año, palabras clave. Este criterio fue creado para atender las necesidades del usuario que busca información en el Repositorio, es decir, que utiliza lenguaje natural, que se refiere al “texto libre” incluido por el usuario en la búsqueda. Al mismo tiempo, se intentó cumplir con un lenguaje controlado, o incluso términos autorizados dentro de la estructura de indexación ya utilizada en el Repositorio.

Se empezó, por tanto, con el tratamiento computacional, utilizando el lenguaje de programación *Python* y la librería *scrapy*. A partir del código desarrollado se pudo crear una función para realizar las operaciones, observando cómo está estructurada la página *web*, y la información dentro de los *strings*. Finalmente, con los datos extraídos, para su tratamiento manual, se instalaron los títulos y su información para que pudieran ser convertidos a un archivo *.xlsx*, donde se almacenaron todos los metadatos previamente definidos (Figura 1). Se extrajeron un total de 1.352 trabajos académicos, entre artículos, libros, tesis, disertaciones, trabajos de conclusión de curso – graduación, informes de investigación. Para este trabajo se hizo un análisis con tesis y disertaciones, resultando en 985 ítems.

Figura 1. Código de extracción de datos en el Repositorio de la UFBA, Fuente:

```
@inline_requests
def parse_Periodico(self, response):
    tableLabels = response.css('.itemDisplayTable').css('td.metadataFieldLabel').extract()
    tableValues = response.css('.itemDisplayTable').css('td.metadataFieldValue').extract()
    tableLabels = [remove_tags(label) for label in tableLabels]
    tableValues = [remove_tags(value) for value in tableValues]
    table = dict(zip(tableLabels, tableValues))

    tipo = table['metadata.dc.type: ']
    titulo = table['Title: ']
    author = table['Authors: ']
    IssueDate = table['Issue Date: ']
    palabrasChave = table['Keywords: ']
    newPalavrasChave = ''

    for i, letter in enumerate(palavrasChave):
        if i and letter.isupper():
            newPalavrasChave += ' '

        newPalavrasChave += letter

    yield {
        'tipo': tipo,
        'titulo': titulo,
        'autores': author,
        'Data': IssueDate,
        'Palavras Chave': newPalavrasChave
    }
```

Elaboración propia.

Con la creación de una base de datos textual, el tratamiento siguió el paso de la minería de textos aplicada a los títulos de trabajos académicos extraídos en el campo de las Humanidades. Cabe señalar que la opción de analizar los títulos de las producciones científicas se basa en el entendimiento de que el lenguaje natural utilizado por los usuarios de un sistema de búsqueda es más recordado cuando se asocia al lenguaje controlado y, en ese sentido, los títulos de las publicaciones incorporan tanto los polos lingüísticos utilizados y además representan temas emergentes y en evolución de un dominio de conocimiento.

Para la formación de la red semántica se tuvo en cuenta la exclusión de las palabras gramaticales, aquellas que sirven para enlazar y organizar palabras en un texto (artículos, preposiciones, conjunciones, pronombres, etc.), manteniendo las palabras léxicas, es decir, que tienen significado y representación dentro de una comunidad. Tal representación se apoyó en el examen de vocabularios controlados en las áreas de Humanidades (*Thesaurus of Indexing Terms*<sup>1</sup>, *UNESCO Thesaurus*<sup>2</sup>, *United Kingdom Archival Thesaurus (UKAT)*<sup>3</sup>, *Tesauro de Folclore e Cultura Popular Brasileira*<sup>4</sup>, *DYAS Humanities Thesaurus*<sup>5</sup>, *TaDIRAH - Taxonomía sobre Actividades de Investigación Digital en Humanidades*<sup>6</sup>, entre otros).

<sup>1</sup>Accesible desde: <https://www.archives.gov/federal-register/cfr/thesaurus.html>

<sup>2</sup>Accesible desde: <https://vocabularies.unesco.org/>

<sup>3</sup>Accesible desde: <https://ukat.aim25.com/thesaurus/>

<sup>4</sup>Accesible desde: <http://www.cnfcp.gov.br/tesauro/>

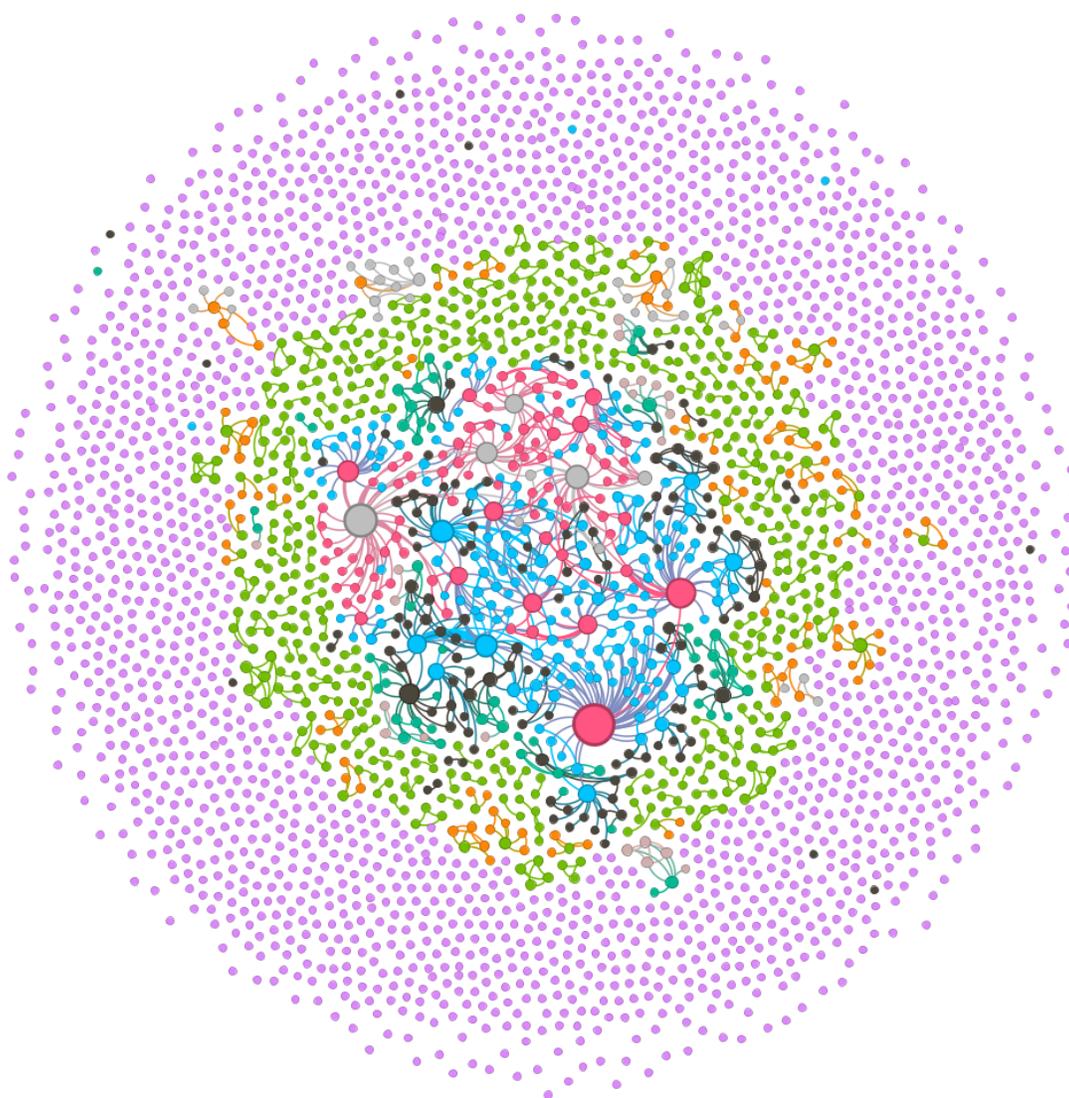
<sup>5</sup>Accesible desde: <https://vocabs.dariah.eu/dyas/en/>

<sup>6</sup>Accesible desde: <https://www.vocabularyserver.com/tadirah/es/index.php>

## Modelación conceptual del campo de las humanidades: una red semántica

Luego de la etapa de minería de texto, fue necesario procesar manualmente los datos textuales, generando 2.819 términos semánticos. Como resultado de esta primera fase de la investigación, se tiene la visualización de conceptos, a partir de la creación de una red semántica, utilizando el *software Gephi*.

Figura 2. Red semántica de términos extraídos de los títulos de tesis y disertaciones en Humanidades de la RI/UFBA



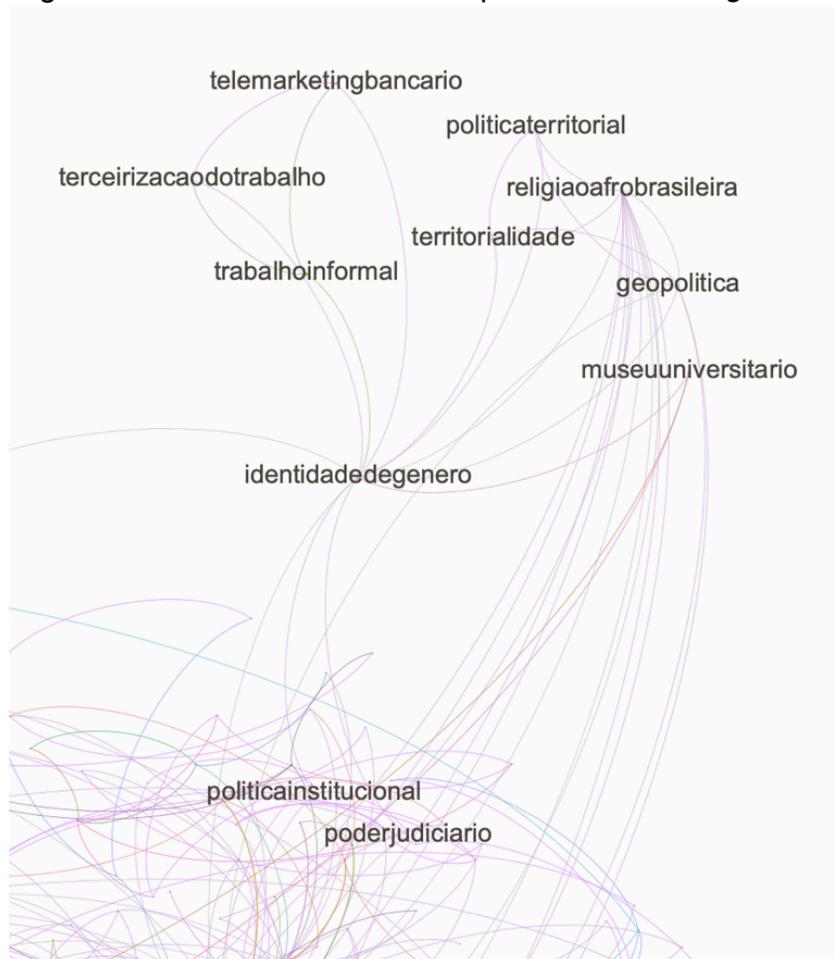
Fuente: Elaboración propia.



una comunidad de conceptos estructurados entre sí, evidenciando conceptos de uso común en la investigación en Humanidades, o tendencias en la investigación en esta área. Otra medida utilizada para caracterizar la red fue el camino promedio más corto entre nodos, que indicaba la distancia entre un concepto y otro (2,83), es decir, la cantidad mínima atribuida a las conexiones que separan un concepto.

Se destaca la comunidad ligada al concepto de *identidad de género*, caracterizada según la relación entre palabras en una misma oración y/o frase, del tipo clique, que considera la vecindad de las palabras para comprender su significado a partir de ellas. el contexto sintagmático en el que está organizado. En este análisis, es posible identificar el núcleo de palabras interconectadas, que componen, a partir de las conexiones, asociaciones entre conceptos extraídos de los títulos de trabajos científicos y su adhesión en el campo de las Humanidades.

Figura 4. Asociaciones del concepto “*identidad de género*”.



Fuente: Elaboración propia.

A partir del esquema estructural de la red de la Figura 4, se categorizaron los términos, organizándolos según objetivaciones pragmáticas a partir de las

relaciones jerárquicas y funcionales entre los conceptos. Como base conceptual se utilizaron vocabularios controlados en Humanidades y literatura académica sobre términos no identificados en estos instrumentos.

Tabla 1. Relaciones entre términos en Humanidades.

Término	Tipo semántico	Relación semántica	
identidad de género	grupos de género	conceptual	Aspecto de la identidad social, que se refiere a las características y roles que asume una persona, tal como lo designan los estereotipos sociales y culturales predominantes sobre su género.
		interactuar con	<i>política institucional, poder judicial; museo universitario; geopolítica; religión afrobrasileña; territorialidad; política territorial, externalización del trabajo; trabajo informal; telemercadeo bancario</i>
política institucional	regímenes políticos	conceptual	Formas de organización y legitimación de la gobernanza, relaciones de poder institucionalizadas y procesos sociopolíticos dentro de un Estado. Los regímenes políticos varían según sus características tipológicas y su contexto histórico.
		interactuar con	<i>identidad de género; poder Judicial</i>
poder Judicial	Sistema judicial	conceptual	El sistema de funciones y actividades realizadas por las autoridades judiciales de un estado u otro organismo soberano para resolver disputas judiciales, hacer frente al crimen y la delincuencia, y

			salvaguardar la aplicación de las leyes.
		interactuar con	<i>identidad de género; política institucional</i>
museo universitario	museos especializados	conceptual	Institución permanente, sin fines de lucro, al servicio de la sociedad, que investiga, colecciona, conserva, interpreta y exhibe el patrimonio material e inmaterial. Los museos, abiertos al público, accesibles e inclusivos, fomentan la diversidad y la sostenibilidad. Los museos funcionan y se comunican de manera ética, profesional y con la participación de las comunidades, y brindan diversas experiencias de educación, fruición, reflexión e intercambio de conocimientos (ICOM, 2022),
		interactuar con	<i>identidad de género</i>
geopolítica	geografía política	conceptual	Tipos de divisiones administrativas utilizadas por un estado u otra estructura soberana o política dentro de los límites de su jurisdicción. Son características intrínsecas, entre otras, la extensión de la superficie sobre la que se aplican/ejerce, el período cronológico en el que están vigentes, sus órganos de gobierno y el grado de autonomía que les confiere en relación

			con su gestión y explotación. de tus recursos.
		interactuar con	<i>política institucional, poder judicial; territorialidad; política territorial</i>
religión afrobrasileña	comunidades religiosas	concepto	Sistema de creencias afrobrasileño con fuerte presencia de la matriz religiosa africana. Los grupos de individuos que tienen la misma fe, dogma y prácticas de culto comparten un sistema común de significados culturales y metafísicos y desarrollan un sentido común de identidad.
		interactuar con	<i>identidad de género; política territorial; geopolítica; territorialidad</i>
territorialidad	derecho humanitario	concepto	El principio de territorialidad es un principio de derecho que permite establecer o delimitar el área geográfica en la que un Estado ejercerá su soberanía.
		interactuar con	<i>geopolítica; religión afrobrasileña; política territorial; identidad de género</i>
política territorial	derecho territorial	concepto	Conjunto de medidas estratégicas de intervención en un territorio, para que asuma las formas propias del conjunto de intereses que controlan el poder político (SANCHEZ, 1992, p. 72)
		interactuar con	<i>geopolítica; religión afrobrasileña;</i>

			<i>territorialidad; identidad de género</i>
telemercadeo bancario	Comercialización	concepto	Estrategia de sucursales bancarias para la comercialización de servicios a través de sistemas de telecomunicaciones.
		interactuar con	<i>identidad de género; subcontratación de trabajo; trabajo informal</i>
trabajo de subcontratación	Directo del trabajo	concepto	Transferencia de la realización de parte del proceso productivo a la responsabilidad de otra empresa.
		interactuar con	<i>identidad de género; telemercadeo bancario; trabajo informal</i>
trabajo informal	Directo del trabajo	concepto	Desarrollo de actividades retribuidas sin vinculación registrada en la tarjeta de trabajo o documentación equivalente.
		interactuar con	<i>identidad de género; telemercadeo bancario; trabajo de subcontratación</i>

Fuente: Elaboración propia.

Con el ejemplo del cuadro anterior, es claro que las relaciones entre términos en el área de Humanidades están relacionadas con cuestiones sociales, sobre todo, enfocadas en temas de raza, género y trabajo, especialmente en trabajos académicos producidos dentro del ámbito de la UFBA-Brasil. En esta sección de la red (clics) se aprecia una tendencia hacia temas humanísticos en la producción científica analizada. Además, aun con el uso, en algunos casos, del lenguaje natural en los títulos de los trabajos académicos, la modelación conceptual realizada permite ampliar el alcance de la recuperación de información sobre el tema, a partir de interacciones con otros términos, formando comunidades.

El análisis conceptual y estructural de las relaciones entre los términos extraídos de las obras académicas de la RI/UFBA, teniendo como punto de partida el lenguaje natural presente en los títulos de estas producciones académicas, contribuye al control del vocabulario en estos sistemas de información con miras a representar con mayor amplitud los contenidos informativos almacenados.

La red semántica creada, como modelo de relación entre términos/conceptos, se basa en un enfoque centrado en el usuario. Caracterizado como un sistema de organización del conocimiento, destaca un enfoque que favorece el comportamiento del usuario en la búsqueda de información en entornos *web*.

## Conclusiones

Con base en el Repositorio Institucional (RI) de la UFBA, esta investigación presenta un modelo conceptual, como metodología para el desarrollo de vocabulario controlado para las áreas de Humanidades. El Proyecto, financiado con becas de iniciación científica de las agencias brasileñas FAPESB y CNPQ, se configura como una investigación transversal, ya que el estudio de las redes conceptuales y semánticas está relacionado con las áreas de Ciencias de la Información, Biblioteconomía y Ciencias de la Computación, ya sea en de forma teórica, práctica o aplicada.

Con la generación de 2.819 términos semánticos, extraídos de trabajos académicos en las áreas de Humanidades, en RI (UFBA), el resultado de esta fase de la investigación, además de visualizar las relaciones entre conceptos, fue la posibilidad de ampliar el alcance de recuperación de información sobre los temas identificados en las encuestas, a partir de interacciones con otros términos, formando comunidades.

Se estima que tal representación de la información científica producida en Humanidades en la UFBA, contribuye como procedimiento metodológico para la creación de un vocabulario controlado especializado. También se espera que, a partir de otros tratamientos descriptivos y temáticos, la interconexión temática entre estos conceptos de distintas áreas del conocimiento, en el ámbito de las Ciencias Humanas, contribuya al alcance social y político de los trabajos depositados en las RI, permitiendo una mirada más refinada a la investigación en Humanidades.

## Bibliografía

- Albert, Réka, e Albert-László Barabási. "Statistical Mechanics of Complex Networks". *Reviews of Modern Physics* 74, n.º 1 (enero de 2002): 47–97. <https://doi.org/10.1103/revmodphys.74.47>.
- Brachman, Ronald J. "Taxonomy, Descriptions, and Individuals in Natural Language Understanding." *ACLWeb. La Jolla, California, USA: Association for Computational Linguistics* 79, (junio de 1979): 33-37. <https://doi.org/10.3115/982163.982174>.
- Café, Lígia, e Marisa Bräscher. "Organização do Conhecimento: Teorias semânticas como base para estudo e representação de conceitos". *Informação & Informação* 16, n.º 2 (15 de diciembre de 2011): 25. <https://doi.org/10.5433/1981-8920.2011v16n2p25>.
- Campos, Maria Luiza de Almeida. "Modelização de domínios de conhecimento: Uma investigação de princípios fundamentais". *Ciência da Informação* 33, n.º 1 (abril de 2004): 22–32. <https://doi.org/10.1590/s0100-19652004000100003>.

- Dahlberg, Ingetraut. "Teoria Do Conceito". *Ciência Da Informação* 7, n.º 2 (1978): 101-107. <https://doi.org/10.18225/ci.inf.v7i2.115>.
- Davis, Hugh, Wendy Hall, Ian Heath, Gary Hill, and Rob Wilkins. "Towards an Integrated Information Environment with Open Hypermedia Systems." *Proceedings of the ACM Conference on Hypertext* 92, (noviembre de 1992):181-190. <https://doi.org/10.1145/168466.168522>.
- Freeman, Linton C. "Centrality in Social Networks Conceptual Clarification". *Social Networks* 1, n.º 3 (enero de 1978): 215–39. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7).
- Lima, Gercina Ângela de. "Organização E representação Do Conhecimento E Da informação Na Web: Teorias E técnicas". *Perspectivas Em Ciência Da Informação* 25, número especial (febrero de 2020): 57-97. <https://periodicos.ufmg.br/index.php/pci/article/view/22283>.
- Newell, Allen. "The Knowledge Level". *Artificial Intelligence* 18, n.º 1 (enero de 1982): 87–127. [https://doi.org/10.1016/0004-3702\(82\)90012-1](https://doi.org/10.1016/0004-3702(82)90012-1).
- Newman, M. E. J. *Networks: An Introduction*. Oxford: Oxford University Press, 2010.
- Novak, Joseph D., e Alberto J. Cañas. "The Development and Evolution of the Concept Mapping Tool Leading to a New Model for Mathematics Education". In *Concept Mapping in Mathematics*, 3–16. Boston, MA: Springer US, 2009. [https://doi.org/10.1007/978-0-387-89194-1\\_1](https://doi.org/10.1007/978-0-387-89194-1_1).
- Novo, Hildenise Ferreira, and José Gacia Vivas Miranda. "Organização Do Conhecimento Na Perspectiva Do Modelo Dinâmico de Análise Conceitual." *Tendências Da Pesquisa Brasileira Em Ciência Da Informação* 8, n.º 2 (2015): 37–58. <https://revistas.ancib.org/index.php/tpbci/article/view/350>.
- Souza, Renato Rocha, Douglas Tudhope and Maurício Barcellos Almeida. "Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems". *Knowledge Organization* 39, n.º 3 (2012): 179–92. <https://doi.org/10.5771/0943-7444-2012-3-179>.
- Strogatz, Steven H. "Exploring Complex Networks". *Nature* 410, n.º 6825 (março de 2001): 268–76. <https://doi.org/10.1038/35065725>.
- Wasserman, Stanley. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1998.
- Watts, Duncan J., e Steven H. Strogatz. "Collective Dynamics of 'Small-World' Networks". *Nature* 393, n.º 6684 (junio de 1998): 440–42. <https://doi.org/10.1038/30918>.
- Watts, Duncan J. "Networks, Dynamics, and the Small-World Phenomenon". *American Journal of Sociology* 105, n.º 2 (septiembre de 1999): 493–527. <https://doi.org/10.1086/210318>.
- Woods, William A. "WHAT'S IN a LINK: Foundations for Semantic Networks". In *Readings in Cognitive Science*, 102–25. Elsevier, 1988. <https://doi.org/10.1016/b978-1-4832-1446-7.50014-5>.

Licencia Creative Commons Attribution  
 Nom-Comercial 3.0 Unported (CC  
 BY-NC 3.0) Licencia Internacional



**CUADERNOS DE SOFÍA**  
**EDITORIAL**

Las opiniones, análisis y conclusiones del autor son de su responsabilidad y no necesariamente reflejan el pensamiento de la Revista.